

Ossi Rintanen

# YRITYKSEN MASTER DATAN LAADUN KARTOITTAMINEN TIETOJÄRJESTEL- MÄUUDISTUKSEN YHTEYDESSÄ

Diplomityö  
Tekniikan ja luonnontieteiden tiedekunta  
Tarkastaja: Henri Pirkkalainen  
Tarkastaja: Samuli Pekkola  
Helmikuu 2020

# TIIVISTELMÄ

Ossi Rintanen: Yrityksen master datan laadun kartoittaminen tietojärjestelmä uudistuksen yhteydessä

Diplomityö

Tampereen yliopisto

Tietojohtamisen tutkinto-ohjelma

Helmikuu 2020

Tietojärjestelmät ja niissä käsiteltävä data ovat merkittävässä roolissa nykypäivän organisaatioiden toiminnassa. Esimerkiksi muuttuneiden tarpeiden tai kasvavan toiminnan seurauksena organisaatiot joutuvat tekemään tietojärjestelmä uudistuksia, joissa uusi järjestelmä korvaa vanhan järjestelmän ja organisaation master data siirretään uuteen järjestelmään. Merkittävän suuri osa näistä tietojärjestelmä uudistuksista epäonnistuu tai kohtaa suuria ongelmia. Ongelmien syitä on monia, mutta karkeasti ne voidaan jakaa prosessi-, käyttäjä- sekä datalähtöisiin ongelmiin. Kirjallisuudessa on käsitelty hyvin rajallisesti sitä, millaisia datalähtöisiä ongelmia tietojärjestelmä uudistuksissa esiintyy ja miten niihin on reagoitava. Tässä työssä lähdettiin selvittämään, miten master datan laadun kartoitus voidaan toteuttaa tietojärjestelmä uudistuksen yhteydessä.

Kirjallisuudessa datan laatuun vaikuttavia tekijöitä on erittäin suuri määrä. Työssä näistä tekijöistä tunnistettiin viisi keskeisintä tekijää: tarkkuus, saatavuus, jatkuvuus, luotettavuus ja turvallisuus. Näille tekijöille tunnistettiin neljä erilaista näkökulmaa: data, tietojärjestelmä, käyttäjä sekä prosessi. Viidestä keskeisestä tekijästä ja neljästä näkökulmasta muodostettiin datan laadun viitekehys, joka kattaa laajasti ja selkeästi datan laadun eri aspektit ja laadun mittaamisen keinot. Tämä viitekehys helpotti datan laadun kokonaisvaltaista tarkastelua.

Työn kohdeyritys on valmistavan teollisuuden pk-yritys, jossa otettiin käyttöön PLM- ja ERP-järjestelmät. Yrityksen master data siirrettiin PLM-järjestelmään, josta se replikoituu ERP-järjestelmään. Työ keskittyi tarkastelemaan PLM-järjestelmään siirrettyä tuotetdataa.

Työssä toteutettiin tapaustutkimus ja empiirinen aineisto kerättiin haastatteluilla. Haastattelulla kohdeyrityksessä datan tuottajia, käyttäjiä sekä hallinnoija tunnistettiin 19 kohdeyrityksen datan laatuun vaikuttavaa ongelmaa. Näihin ongelmiin muodostettiin toimenpidesuosituksot luokittelemalla ongelmat neljän näkökulman mukaisesti data-, tietojärjestelmä-, käyttäjä- sekä prosessiongelmiin. Dataan liittyvät ongelmat on korjattava ennen datamigraatiota, sillä nämä ongelmat siirtyvät datan mukana uuteen järjestelmään. Tietojärjestelmä ongelmat voidaan katsoa tietojärjestelmä uudistuksen yhdeksi syyksi ja lähtökohtaisesti ne korjaantuvat uudistuksessa. Käyttäjä ongelmat todennäköisesti muuttuvat uuden tietojärjestelmän myötä, mutta käyttäjä ongelmia on ennakoitava ja ennaltaehkäistävä esimerkiksi käyttäjä koulutuksella. Prosessiongelmat puolestaan edellyttävät dataan liittyvien prosessien tarkkaa määrittelyä, sillä prosessinäkökulma asettaa raamit muille datan laadun näkökulmille.

Työn keskeinen huomio datan laadusta on sen laaja-alainen vaikutus koko organisaatioon sekä kontekstisidonnaisuus. Datan laatua on aina tarkasteltava kontekstissa. Kontekstin muuttuessa, eli esimerkiksi tietojärjestelmä uudistuksessa, datan laatuun on kiinnitettävä erityistä huomiota. Toisaalta on myös tiedostettava, että datan laadun huomioiminen ei ole yksittäinen toimenpide, vaan jatkuvaa työtä.

Avainsanat: tietojärjestelmä, tietojärjestelmä uudistus, master data, datan laatu, datamigraatio, ERP, PLM.

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

# ABSTRACT

Ossi Rintanen: Mapping master data quality in organization as part of information system implementation.

Master of Science Thesis

Tampere University

Information and Knowledge Management

February 2020

---

Information systems and data inside them are in a crucial role in today's organizations. For example, due to changing requirements and growing scale of operations organizations have to upgrade their information systems. In these upgrades an old information system can be replaced by a new one, which also requires a migration of master data from the old system to the new one. A significant portion of these information system upgrades fail or experience major problems. There are many causes for these problems but roughly they can be categorized to process, user and data related problems. In literature there is very limited amount of information about data related problems and appropriate reactions to them. This thesis aims to answer the question of how to execute a mapping of master data quality as part of information system implementation.

There are a great number of data quality factors in literature. In this thesis five of these were recognized as most central factors: accuracy, availability, continuity, reliability and security. For these factors four different perspectives were recognized: data, information system, user and process. These five central quality factors and four perspectives formed a data quality framework which covers widely and distinctly different aspects and measuring of data quality. This framework was used to examine data quality comprehensively.

The company studied in this thesis is a medium sized enterprise of manufacturing industry. The company implemented PLM and ERP systems. The master data of the company was migrated to the PLM system from which it was replicated to the ERP system. The research of this thesis concentrated on the product data that was migrated to the PLM system.

A case study was executed for this thesis. Empirical material was gathered by interviewing data producers, users and administrators in the company. These interviews revealed 19 problems affecting the data of the company. Action recommendations for each problem was formed by categorizing the problems according to the four data quality perspectives to data, information system, user and process related problems. Data problems must be fixed before migration as these problems move with the data to the new system. Information system related problems can be considered as a reason for implementing a new system and thus should disappear in implementation. User related problems most likely change due to the implementation of the new system, but user problems should be anticipated and preempted with, for example, user training. Process related problems require precise defining of data related processes, because process perspective sets the frame for other data quality perspectives.

The most important finding of the study was the companywide impact of data quality and its tight relation to context. Data quality must always be assessed in a context. When context is changed, for example when implementing a new information system, data quality must be reassessed. On the other hand, it should be acknowledged that assessing data quality is not a single but continuous operation.

Keywords: information system, information system implementation, master data, data quality, data migration, ERP, PLM

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

# ALKUSANAT

Diplomityö vaikutti jo etukäteen suurelta työltä, mutta työn pitkittynyt aikajänne tuli silti yllätyksenä. Jo pelkästään aiheen muotoutumiseen meni jokunen kuukausi. Lopulta työ venyi peräti kahdelle vuosikymmenelle. Jos saisin jälkikäteen päättää, en ehkä tekisi diplomityötä töiden ohessa, vaan varaisin sille oman aikansa.

Haluan kiittää erittäin paljon ohjaajiani Henri Pirkkalaista ja Samuli Pekkola erinomaisesta ja rakentavasta palautteesta. Paljon kiitoksia myös kohdeyrityksen yhteyshenkilöille aktiivisesta ja kiinnostuneesta otteesta työtä kohtaan (sekä kärsivällisyydestä). Kiitokset palautteesta myös Risto Silvolalle. Lisäksi maltillinen kiitos kaikille, jotka toistuvasti kyselivät, miten diplomityö edistyy, mahdollisesti näin vauhdittaen osaltaan työn valmistumista.

Tampereella, 4.2.2020

Ossi Rintanen

# SISÄLLYSLUETTELO

|   |    |
|---|----|
| 1. JOHDANTO .....   | 1  |
| 1.1 Tietojärjestelmä uudistukset ja datamigraatiot.....                 | 2  |
| 1.2 Tutkimuskysymys.....  | 3  |
| 2. TIETOJÄRJESTELMÄUUDISTUKSET JA MASTER DATA .....                     | 5  |
| 2.1 Tietojärjestelmät organisaatioissa .....                            | 5  |
| 2.2 Tietojärjestelmä uudistukset.....                                   | 6  |
| 2.3 Tietojärjestelmän sopivuus.....                                     | 8  |
| 2.4 Datan kategoriat.....   | 10 |
| 2.5 Master data ja sen ominaispiirteet.....                             | 11 |
| 3. DATAN LAADUN MÄÄRITELMÄ JA VIITEKEHYS .....                          | 13 |
| 3.1 Datan laadun osatekijät kirjallisuudessa.....                       | 13 |
| 3.2 Datan laadun viitekehys .....                                       | 15 |
| 3.3 Datan laadun osatekijöiden kuvaukset .....                          | 17 |
| 4. MASTER DATAN LAADUN HALLINTA.....                                    | 26 |
| 4.1 Datan laadunhallintaprosessimalli .....                             | 26 |
| 4.2 Datan laadun näkökulmien ominaispiirteet .....                      | 28 |
| 5. TUTKIMUSMENETELMÄ .....  | 35 |
| 5.1 Tapaustutkimus.....   | 35 |
| 5.2 Kohdeyritys .....   | 36 |
| 5.3 Empiirisen aineiston kerääminen.....                                | 37 |
| 5.4 Empiirisen aineiston analysointi .....                              | 39 |
| 6. TULOKSET .....   | 40 |
| 6.1 Näkemys datan laadusta ja siihen suhtautuminen.....                 | 40 |
| 6.2 Datan käyttö ja datan laatu ennen tietojärjestelmä uudistusta ..... | 44 |
| 6.3 Datan laatu nykyhetkellä .....                                      | 51 |
| 6.4 Datan laatu tulevaisuudessa .....                                   | 55 |
| 7. POHDINTA JA JOHTOPÄÄTÖKSET .....                                     | 65 |
| 7.1 Havaittujen ongelmien luokittelu.....                               | 65 |
| 7.2 Yleiset havainnot tutkimuksesta .....                               | 72 |
| 7.3 Teoreettiset kontribuutiot.....                                     | 73 |
| 7.4 Käytännön kontribuutiot .....                                       | 74 |
| 7.5 Tutkimuksen rajoitteet.....   | 75 |
| 7.6 Tulevaisuuden tutkimusaiheet.....                                   | 76 |
| 7.7 Loppusanat .....  | 77 |
| LÄHTEET .....   | 79 |

|  |    |
|--|----|
| LIITE A: HAASTATTELUTUTKIMUKSEN KYSYMYSRUNKO ..... | 84 |
|--|----|

## LYHENTEET JA MERKINNÄT

|               |  |
|---------------|--|
| CAD           | engl. Computer Aided Design, tietokoneavusteinen suunnittelu                         |
| CRM           | engl. Customer Relationship Management, asiakkuuksien hallinta                       |
| Datamigraatio | Prosessi, jossa data, useimmiten master data, siirretään yhdestä järjestelmä toiseen |
| ERP           | engl. Enterprise Resource Planning, toiminnan- ja resurssien ohjaus                  |
| MDM           | engl. Master Data Management, master datan hallinta                                  |
| PLM           | engl. Product Lifecycle Management, tuotteen elinkaaren hallinta                     |

# 1. JOHDANTO

Organisaatiot tuottavat ja käyttävät dataa monilla eri tavoilla. Kaikissa operatiivisissa päätöksissä hyödynnetään dataa (Silvola 2018), joten päätökset ovat yhtä laadukkaita kuin niiden pohjana käytetty data (Redman 1998). Lisäksi dataa käytetään tuotteiden, palveluiden ja tukitoimintojen laadun varmistamiseen (Silvola 2018). Dataa voidaankin sanoa käytettävän organisaatioiden arvonluontiketjujen varhaisissa vaiheissa, jolloin datan ongelmat heijastuvat myös arvonluontiketjujen myöhempiin vaiheisiin.

Liiketoiminnan ja datan ongelmat liittyvät oleellisesti toisiinsa. Lee et al. (2014) nostavatkin esille, että monesti dataan liittyvien ongelmien syyt johtuvat liiketoiminnan ongelmista. Kirjallisuudessa on laajalti korostettu IT- ja liiketoimintastrategioiden yhdistämisen merkitystä (Tallon 2008, Preston & Karahanna 2009, Yayla & Hu 2012, Gerow et al. 2015, Reynolds & Yeatton 2015, Silvola 2018). Datan, erityisesti master datan, laatu kuvaa miten tämä strategioiden yhdistäminen toimii käytännössä (Silvola 2018). Liiketoimintastrategian lisäksi datan hallinta on organisoitava yhteen myös tuotehallinnan kanssa, mistä on tullut yksi yritysten olemassaolon keskeinen vaatimus (Buffington 2011, Ouertani et al. 2011).

Datan laadun ongelmat ovat todella yleisiä. Tietojärjestelmien kehittyminen mahdollistaa entistä suurempien ja kompleksisempien datamassojen keräämisen (Watts et al. 2009, Haug & Stentoft 2011). Tämä asettaa alati kasvavat paineet datan hallinnalle. Datan laadun ongelmat johtuvatkin siitä, että datan hallintaan ei ole panostettu riittävästi huolimatta datan määrän kasvusta (Lee et al. 2006, Knolmayer & Röthlin 2006, Breuer 2009). Silvola (2018) huomauttaa dataan liittyvistä haasteista tulleen arkipäivää, koska datan laatuun ei ole panostettu aiemmin. Erään tutkimuksen mukaan yli 60 % keskikokoisista yrityksistä koki ongelmia datan laadun kanssa (Wand & Wang 1996).

Datan heikolla laadulla on vakavia vaikutuksia organisaation toiminnan tehokkuuteen (Wand & Wang 1996). Datan laatuongelmat hankaloittavat päätöksentekoa ja lisäävät operatiivisia kustannuksia, koska resursseja joudutaan käyttämään ongelmien löytämiseen ja korjaamiseen (Redman 1998). Arvioiden mukaan organisaatiot menettävät 15-25 % käyttökatteestaan (Olson 2003) tai 8-12 % liikevaihdostaan datan laadun ongelmien vuoksi (Redman 1998). On myös arvioitu, että Yhdysvalloissa heikkolaatuinen



asiakasdata maksaa yhdysvaltalaisille yrityksille vuodessa \$611 miljardia (Foley & Helfert 2010). Datan laatuongelmat heikentävät myös asiakastytyvääisyyttä, työntekijöiden moraalialia sekä uskoa organisaatioon (Redman 1998, Russom 2006).

## 1.1 Tietojärjestelmäüudistukset ja datamigraatiot

Säilyttääkseen kilpailukykynsä muuttuvassa liiketoimintaympäristössä, organisaatioiden on pakko uudistaa toimintojaan ja käytäntöjään (Bhatti 2014). Lisäksi organisaatioiden eläessä ja kasvaessa niiden tarpeet datan ja tietojärjestelmien suhteen muuttuvat. Korkealaatuisen datan tuottaminen ja ylläpitäminen vaativat sekä datan että dataa käsittelevien tietojärjestelmien jatkuvaa mukautumista muuttuvien markkinoiden aiheuttamiin tarpeisiin (Strong et al. 1997). Työn kohdeyrityksessä toteutettiin tietojärjestelmäüudistus, joka alun perin nosti esille kysymyksen datan laadun parantamisesta ja ylläpitämisestä. Hyvälaatuinen data on yksi keskeinen onnistumistekijä tietojärjestelmäüudistuksissa (Hong & Kim 2002). Uusien tietojärjestelmien käyttöönnotto vaatii usein vanhan datan siirtämisen uuteen ympäristöön eli datamigraation. Datan, erityisesti master datan, laatu on tärkeää huomioida datamigraation ja konversioprojektien yhteydessä.

Vaikka tietojärjestelmät ja niiden sisältämä data ovat nykypäivän organisaatioille kriittisiä, valitettavan harva organisaatio pyrkii mittaamaan käyttämänsä datan laatua. Lisäksi noin puolilla yrityksistä ei ole mitään suunnitelmaa datan laadun parantamiseksi. (Foley & Helfert 2010) Monet uskovat, että tietojärjestelmät hoitavat dataan liittyvät ongelmat automaattisesti (Redman 2008). Uudesta järjestelmästä saatu hyöty vähenee merkittävästi, jos vanhat ongelmat vain siirretään uuteen järjestelmään. Lisäksi datamigraation ja konversioprojektien katsotaan itsessään olevan kolmanneksi yleisin syy datan laadullisille ongelmille (Russom 2006).

Työ tarkastelee datan laatua tietojärjestelmäüudistuksen kontekstissa. Kuten myöhemmin todetaan, datan laatu näyttölee merkittävää osaa uusien tietojärjestelmien käyttöönnoton sujuvuuden ja onnistumisen kannalta. Uudistusvaihetta ja tietojärjestelmien käyttöönnottoa hyödynnetään ajallisena käännekohtana, jossa pystytään korjaamaan datan vanhoja laatuongelmia sekä ottamaan käyttöön uusia datan luomiseen, käyttöön ja ylläpitoon liittyviä käytäntöjä. Tietojärjestelmäüudistusta ja datamigraatiota voidaan verrata asunnon vaihtamiseen, jossa uusi konteksti näyttää vanhat tavarat uudessa valossa. Samassa yhteydessä saattaa myös paljastua, mikä on turhaa ja ylimääräistä ja missä on puutteita.

Tietojärjestelmiä myös kehitetään työn kirjoitushetkellä jatkuvasti. Korkealaatuista dataa tuottavien tietojärjestelmien ja prosessien suunnittelu edellyttää datan laadun hyvää tuntemusta (Wand & Wang 1996), minkä vuoksi työ on erityisen ajankohtainen kohdeyritykselle.

## 1.2 Tutkimuskysymys

Kaikki edellä mainitut asiat pohjautuvat datan ja sen laadun ymmärtämiseen. Työn pää tavoitteena on luoda organisaatioon käsitys siitä, mitä on hyvälaatuinen data, edistää osaltaan kohdeyrityksen master datan laatua sekä muodostaa pohja kohdeyrityksen tulevalle datan laadunhallintaprosessille uudessa tietojärjestelmäkontekstissa.

Työn tutkimuskysymys on:

Miten kohdeyrityksen master datan laadullisia ongelmia voidaan kartoittaa tietojärjestelmä uudistuksen yhteydessä ja miten niihin on reagoitava?

Tähän kysymykseen vastaamiseksi muodostettiin apukysymykset:

1. Mikä rooli datan laadulla on tietojärjestelmä uudistuksissa ja mitä on master data?
2. Mistä master datan laatu koostuu?
3. Miten master datan laatua mitataan?

Data voidaan jakaa moniin eri luokkiin, joista yksi on master data. Datan luokittelu ja raja-  
aus on esitetty luvussa 2. Luku 2 vastaa ensimmäiseen apukysymykseen, mikä rooli datan laadulla on tietojärjestelmä uudistuksissa ja mitä on master data. Pelkkä data ilman kontekstia ei kuitenkaan ilmennä kuin vain osan laadustaan. Datan laatua on todella hankalaa tarkastella esimerkiksi ilman tietojärjestelmä kontekstia, mikä havaitaan datan laadun osatekijöitä käsittelevässä luvussa 3. Luvussa 3 vastataan myös toiseen apukysymykseen, mistä datan laatu koostuu. Luvussa 4 kuvataan kirjallisuuden pohjalta datan kehittämisprosessimalli ja käsitellään datan laadun mittaamista. Luvussa 4 vastataan kolmanteen apukysymykseen, miten datan laatua mitataan.

Koska datan laatu on osittain sidoksissa muun muassa tietojärjestelmiin ja tietojärjestelmät ovat usein kullekin yritykselle erikseen räätälöityjä, datan laadun tarkastelu on aina yrityskohtaista, mikä tukee tapaustutkimuksen käyttöä. Kysymykseen datan laadun kar-

toittamisesta voitaisiin periaatteessa vastata yleisellä tasolla kirjallisuustutkimuksen perusteella, mutta tällöin menetettäisiin kohdeyritykselle tyypilliset piirteet. Tutkimusmenetelmänä työssä hyödynnetään haastattelututkimusta. Tapaustutkimus ja tutkimusmenetelmät on esitelty luvussa 5. Luvussa 6 on esitetty tutkimuksen tulokset. Luvussa 7 tutkimustuloksia pohditaan aiemmin työssä esitetyn teorian pohjalta.

## 2. TIETOJÄRJESTELMÄUUDISTUKSET JA MASTER DATA

Tässä luvussa perehdytään tarkemmin tietojärjestelmäuudistuksiin sekä datan luokitteluun. Luvussa kerrotaan, miksi datan laatu on tärkeässä osassa tietojärjestelmäuudistuksissa. Lisäksi luvun tarkoituksena on rajata, millaisen datan laatua työssä tarkastellaan.

### 2.1 Tietojärjestelmät organisaatioissa

Tietojärjestelmillä katsotaan olevan potentiaalia parantaa organisaation toimintaa sekä vähentää kustannuksia (Fichman et al. 2011). Tietojärjestelmillä pyritään integroimaan eri osastot ja toiminnot, mikä korostuu erityisesti valmistavassa teollisuudessa (Chrysolouris et al. 2009) ja maantieteellisesti hajautuneissa organisaatioissa (Holland & Light 1999). Tietojärjestelmien nähdään paitsi kokoavan yhteen organisaation osaamista myös lyhentävän suunnitteluun ja kehitykseen käytettävää aikaa ja kustannuksia sekä parantavan yhteistyötä muiden yritysten ja toimittajien kanssa (Chrysolouris et al. 2009).

Yrityksille tärkeimpiä tietojärjestelmiä ovat ERP- ja CRM-järjestelmät sekä PLM-järjestelmä yhdistettynä CAD-ohjelmistoon (Silvola 2018, s30). ERP-järjestelmä on organisaation resurssien kohdistamiseen ja käsittelyyn erikoistunut järjestelmä. ERP-järjestelmillä pyritään parantamaan toimitusketjujen kustannustehokkuutta, lyhentämään läpimenoaikoja, pienentämään varastoja, parantamaan laatua sekä tehostamaan koko organisaation toiminnan tehokkuutta. ERP-järjestelmä integroi yhteen suurien organisaatioiden eri toimipisteet, luoden näin kattavan yleiskuvan koko organisaation toiminnoista ja toimipisteistä. Lisäksi se pitää kirjaa kaikista organisaation transaktioista. (Umble et al. 2003) CRM- eli asiakkuudenhallintajärjestelmällä (*engl. Customer Relationship Management*) puolestaan ylläpidetään organisaation kontaktien tietoja, kuten asiakkaiden yhteystietoja ja tilaushistoriaa. CRM on usein integroitu osaksi ERP-järjestelmää. (Silvola 2018, s36-37)

Tuotteen elinkaaren hallinta (*engl. Product Lifecycle Management, PLM*) on prosessi, joka sisältää tuotteen suunnittelun, valmistuksen, huollon sekä tuotteen hävittämisen. Tämä prosessi yhdistää tietoa, ihmisiä, muita prosesseja sekä liiketoimintajärjestelmiä tarjoten yrityksen tuoteinformaatiolle perusrungon. (Martio 2015, s9) PLM-järjestelmän tarkoitus on integroida tämä prosessi sekä mahdollistaa sen hallinta (Chrysolouris et al.

2009). PLM-järjestelmä on tuotteen elinkaaren hallintaan tarkoitettu, usein yritykselle räätälöity, tietojärjestelmä, joka sisältää sekä geneerisen että tuoteyksilöiden tuotetiedon niiden koko elinkaaren ajalta (Martio 2015, s9). Järjestelmä mahdollistaa esimerkiksi tuoterakenteen tehokkaan käsittelyn sekä muutoshallinnan (Chryssolouris et al. 2009).

Tyypillinen PLM-järjestelmän käyttäjä tekee työtään jollakin dokumenttityökalulla, kuten CAD-ohjelmalla (*engl. Computer Aided Design*). CAD-ohjelmistolla tuotetaan graafisia esityksiä tuotteista. Tätä toiminnallisuutta hyödynnetään erityisesti tuotekehityksessä (Silvola 2018, s37). Martio (2015, s100) huomauttaa, että dokumenttien sisään- ja uloskirjaukset ovat ”ylimääräisiä” tehtäviä, jotka pitäisi pystyä toteuttamaan mahdollisimman yksinkertaisesti. Tästä syystä CAD-ohjelmisto on usein integroitu PLM-järjestelmään.

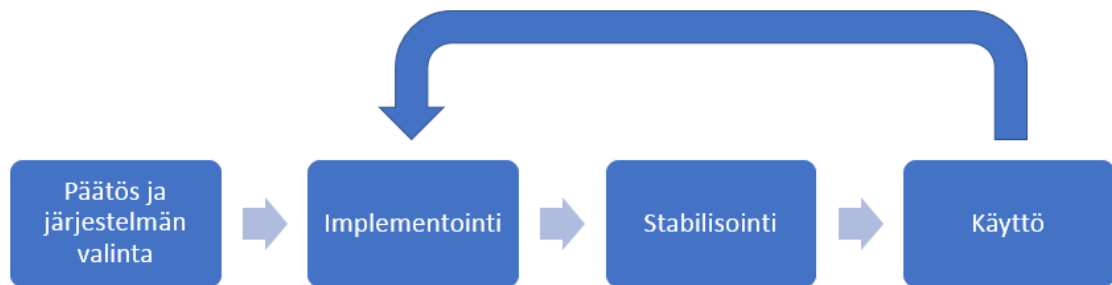
Tämä työ keskittyy tarkastelemaan PLM- ja ERP-järjestelmiä, sekä sivuaa CAD-ohjelmistoa, sillä nämä ovat kohdeyritykselle keskeisimmät järjestelmät. CRM on kohdeyrityksessä toteutettu osana ERP-järjestelmää.

## 2.2 Tietojärjestelmä uudistukset

Tietojärjestelmä uudistuksessa organisaatio ottaa käyttöön uuden tietojärjestelmän. Tällaisiin tietojärjestelmä uudistuksiin liittyy korkea epäonnistumisprosentti (Hong & Kim 2002, Chen et al. 2009). Tietojärjestelmä uudistukset asettavat organisaatiolle usein valtavat paineet ajan ja resurssien suhteen (Bhatti 2014). Tietojärjestelmien, kuten ERP-järjestelmän, tärkeydestä johtuen epäonnistuminen tietojärjestelmä uudistuksessa saattaa vaarantaa jopa organisaation ydintoiminnot (Hong & Kim 2002).

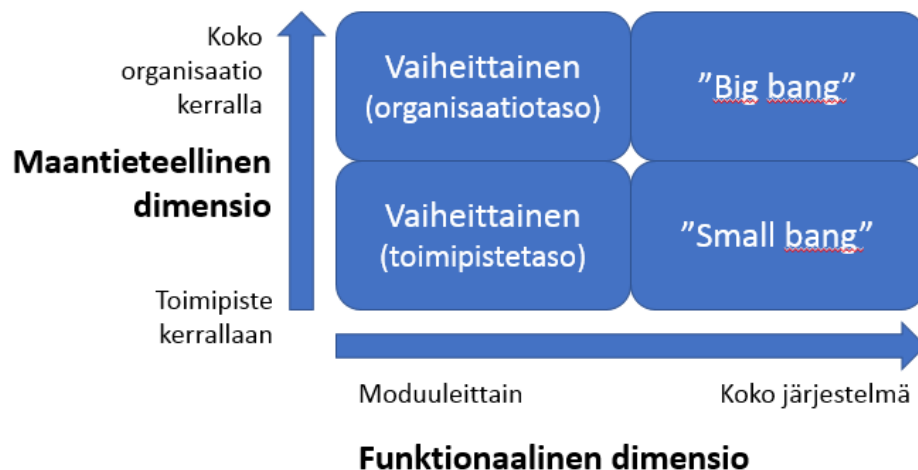
Zwicker & de Souza (2005) kuvaavat tietojärjestelmän käyttöönoton tapahtuvan neljässä vaiheessa. Ensimmäinen vaihe on päätös tietojärjestelmä uudistuksesta ja järjestelmän valinta. Tässä vaiheessa myös valitaan järjestelmän toimittaja sekä suunnitellaan implementointi, joka on prosessin toinen vaihe. Implementointivaihe pitää sisällään uuden järjestelmän kustomointia ja konfigurointia, datamigraation, käyttäjien koulutusta ja lopulta järjestelmän käyttöönoton. Käyttöönoton kolmas vaihe, jota Zwicker & de Souza (2005) kuvaavat erityisen kriittiseksi vaiheeksi tietojärjestelmä uudistuksen onnistumisen kannalta, on stabilisointi. Tämä vaihe vaatii paljon hallinnollista tukea ja järjestelmän seuraamista, sillä tässä vaiheessa esiin nousevat implementointivaiheessa huomaamatta jääneet operationaaliset ongelmat, järjestelmän bugit sekä käyttäjien koulutuspuutteet. Prosessin neljäs vaihe on käyttövaihe, jossa uudesta tietojärjestelmästä on tullut oleellinen ja päivittäinen osa organisaation toimintaa. Tässä vaiheessa huomiota tulee kiinnit-

tää käyttäjien palautteeseen ja uusiin kehitysideoihin, jotka jälleen voidaan viedä implementointivaiheeseen. (Zwicker & de Souza 2005) Kuvassa 2.1 on esitetty Zwicker & de Souzan (2005) tietojärjestelmän käyttöönottoprosessi.



**Kuva 2.1.** Tietojärjestelmän käyttöönottoprosessi. (mukaillen lähteestä Zwicker & de Souza 2005)

Prosessin toinen vaihe eli implementointi voidaan toteuttaa monella eri tavalla. Zwicker & de Souza (2005) esittävät tähän kaksi dimensiota: funktionaalisen ja maantieteellisen dimension. Funktionaalinen dimensio kuvaa sitä, implementoidaanko tietojärjestelmä osissa eli moduuleissa vai kokonaan. Maantieteellinen dimensio puolestaan kertoo, tapahtuuko implementointi kaikkiin organisaation toimipisteisiin kerralla vai edetäänkö käyttöönotossa yksi toimipiste kerrallaan. (Zwicker & de Souza 2005) Kuva 2.2 esittää näistä kahdesta dimensiosta muodostettua nelikenttää.



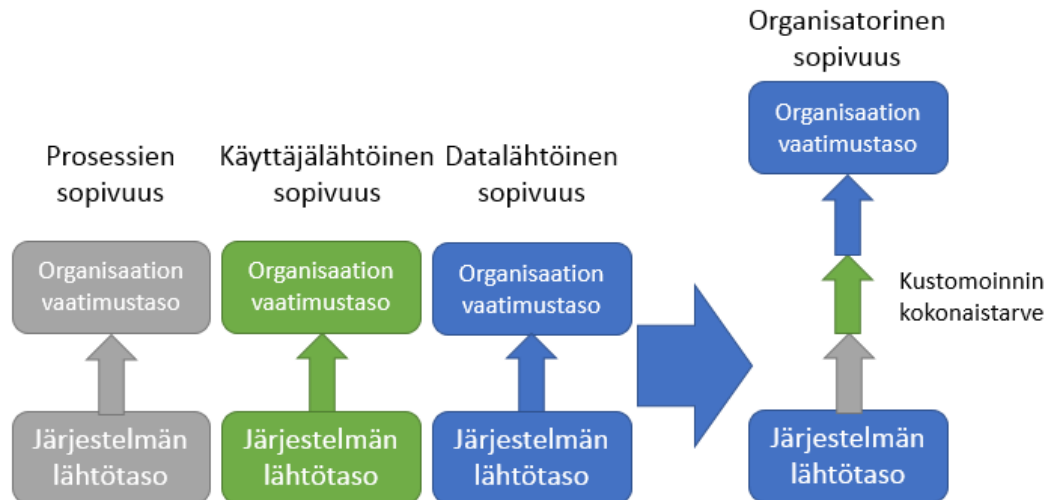
**Kuva 2.2.** Implementoinnin dimensiot. (mukaillen lähteestä Zwicker & de Souza 2005)

Zwicker & de Souza (2005) mukaan siirryttäessä kuvan 2.2 nelikentässä ylös ja oikealle riski organisaation toimintojen keskeytymiseen kasvaa. Suurin riski on ”big bang” implementoinnissa, jossa koko järjestelmä otetaan samaan aikaan kokonaisuudessaan käyttöön kaikissa organisaation toimipisteissä. Tämä lähestymistapa vaatii eniten työtä stabilisaatio vaiheessa. Lisäksi ongelmatilanteissa palaaminen vanhaan järjestelmään on käytännössä mahdotonta. ”Small bang” implementointiin liittyy nämä samat riskit, mutta ne kohdistuvat vain yhteen toimipisteeseen kerrallaan. Vaiheittaisissa implementoinneissa riski koko organisaation tai toimipisteen pysähtymiseen on pienempi ja ongelmatilanteissa paluu vanhaan järjestelmään onnistuu helpommin kuin koko järjestelmän implementoinnissa. Toisaalta vaiheittaisessa implementoinnissa on usein useita päällekkäisiä implementointi ja stabilisointi -vaiheita käynnissä, mikä vähentää yksittäisten vaiheiden saamaa huomiota. (Zwicker & de Souza 2005)

### 2.3 Tietojärjestelmän sopivuus

1990-luvun alusta lähtien organisaatiot ovat kasvavissa määrin siirtyneet kyseiselle organisaatiolle luoduista tietojärjestelmistä ostamaan valmiita tietojärjestelmiä (Holland & Light 1999, Hong & Kim 2002). Etuina valmiissa järjestelmissä ovat pienemmät kustannukset, nopeampi käyttöönotto sekä järjestelmän korkeampi laatu (Hong & Kim 2002).

Ongelmana valmiiden tietojärjestelmien ostamisessa on, että ne harvoin täyttävät sellaisinaan tietyn organisaation tarpeet (Hong & Kim 2002, Chen et al. 2009). Näin ollen valmiina ostettuja tietojärjestelmiä joudutaan kustomoimaan kunkin organisaation tarpeisiin. Hong & Kim (2002) esittävät kullakin valmiilla tietojärjestelmällä olevan tietty organisatorinen sopivuus (*engl. organizational fit*), joka kuvaa tietojärjestelmän kustomointitarpeen määrää. Järjestelmän, jolla on heikko sopivuus, käyttöönotto ja käyttö edellyttää suurta määrää kustomointia, kun taas korkeamman sopivuuden järjestelmän kanssa organisaatio selviää vähemmällä kustomoinnilla. Organisatorinen sopivuus koostuu kolmesta osasta, jotka ovat prosessien sopivuus, käyttäjälähtöinen sopivuus sekä datalähtöinen sopivuus (Hong & Kim 2002). Nämä sopivuudet kuvaavat yleisesti sitä, miten hyvin tietojärjestelmä täyttää organisaation prosessien, henkilöstön ja datan tarpeet. Umble et al. (2003) toteavat, että ERP-järjestelmien tapauksessa useimmissa organisaatioissa olemassa olevat prosessit ja organisaatorakenne eivät vastaa sellaisenaan järjestelmän vaatimuksiin. Kuvassa 2.3 on havainnollistettu organisatorisen sopivuuden ja järjestelmän kustomoinnin kokonaistarpeen muodostumista kolmesta edellä mainitusta tekijästä.



**Kuva 2.3.** Organisatorisen sopivuuden ja järjestelmän kustomoinnin kokonaistarpeen muodostuminen.

Organisatorinen sopivuus on huomioitava tietojärjestelmän implementointivaiheessa. Sekä Esteves & Pastor (2005) että Parr et al. (1999) listaavat yhdeksi implementoinnin kriittiseksi menestystekijäksi kustomoinnin välttämisen. Kustomoinnin välttäminen nähdään jopa merkittävimpänä tekijänä tietojärjestelmän käyttöönottoprojektin budjetti- ja aikataulutavoitteiden saavuttamisen kannalta (Parr et al 1999). Myös Hong & Kim (2002) ja Chen et al. (2009) tunnistavat suuren kustomoinnin johtavan todennäköisemmin implementoinnin epäonnistumiseen. Näin ollen, parantaakseen organisatorista sopivuutta, organisaation on parempi laskea omaa vaatimustasoaan lähemmäs järjestelmän lähtötasoa kuin nostaa järjestelmän lähtötasoa kustomoinnilla organisaation vaatimustasolle.

Tämä työ keskittyy tarkastelemaan näistä kolmesta tekijästä datalähtöistä sopivuutta. Datan laadulla katsotaan olevan merkittävä vaikutus datamigraation eli datan siirtämisen vanhasta järjestelmästä uuteen järjestelmään onnistumisessa sekä lopulta koko tietojärjestelmän käyttöönoton onnistumisessa (Somers & Nelson 2001, Umble et al. 2003). Tästä johtuen datan laadun kartoitus ja datan laatuun panostaminen tietojärjestelmä-uudistuksen yhteydessä on erittäin tärkeää. Datamigraatio toteutetaan uuden tietojärjestelmän käyttöönoton implementointivaiheessa. Datan huolellisella valmistelulla voidaan laskea organisaation datalähtöistä vaatimustasoa lähemmäs tietojärjestelmän lähtötasoa, vähentäen näin kustomointitarvetta.

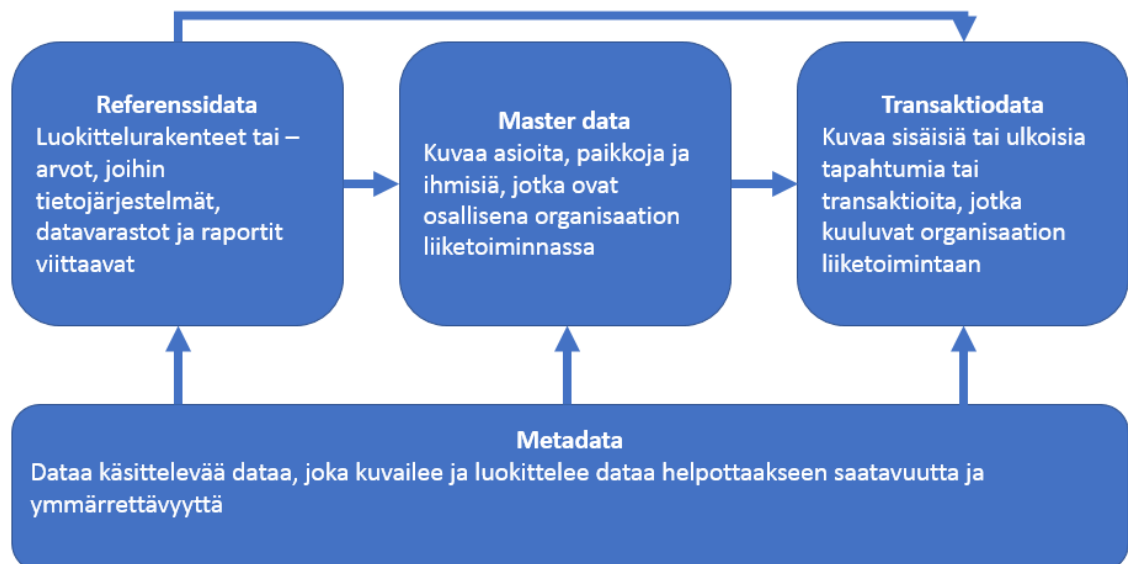
On huomattava, että datan näkökulmasta ”big bang” ja ”small bang” eli koko järjestelmän implementoinnit ovat todella työläitä, koska nämä edellyttävät koko organisaation datan



valmistelua. Vaiheittaisissa implementoinneissa sen sijaan voidaan keskittyä vain implementoitavan moduulin käyttämän datan valmisteluun ja sen migraatioon. Datamigraatiossa siirrettävä data on useimmiten pääasiassa organisaation master dataa, jota käsitellään seuraavaksi.

## 2.4 Datan kategoriat

Data voidaan jakaa erilaisiin kategorioihin, datan luonteenpiirteiden ja ominaisuuksien mukaan (McGilvray 2008, s39). McGilvray (2008) jakaa datan neljään kategoriaan: master data, transaktiodata, referenssidata ja metadata. Nämä kategoriat ja niiden keskeiset ominaispiirteet ja keskinäiset suhteet on esitetty kuvassa 2.4.



**Kuva 2.4.** Datan kategoriat ja niiden väliset suhteet. (mukaillen lähteestä McGilvray 2008, ss. 40-43)

Referenssidataa ovat esimerkiksi statuskoodit, tuotehierarkia ja tuotetyypit, lista-arvot sekä lyhenteet. Referenssidatan standardointi on tärkeä osa organisaation datan hallintaa. Metadataan kuuluu esimerkiksi tietomallia kuvaava data, kuten tietokenttien nimet, tyypit, pituudet ja asettelu sekä tieto siitä, kuka datan on luonut, milloin sitä on muutettu tai milloin se on poistettu. Transaktiodataa ovat mm. myynti- ja ostotilaukset, toimitusdokumentit ja maksut. Master dataan puolestaan kuuluu data, joka kuvaa organisaation liiketoiminnalle keskeisiä asioita, paikkoja ja ihmisiä. (McGilvray 2008, ss. 42-43)

## 2.5 Master data ja sen ominaispiirteet

Master datan voidaan katsoa olevan yrityksen tärkeintä dataa. Master data on yrityksen liiketoiminnan ytimessä olevaa tuotetietoa, asiakastietoa ja toimittajatietoa. Silvola (2018) mainitsee, että master data luodaan tuotekehitysvaiheessa, joka sitten vapautetaan muiden toimintojen käyttöön. Master data on usein validoitu monessa vaiheessa ja sitä säilytetään useimmiten PLM-järjestelmässä (Silvola 2018). Master data voidaan edelleen jakaa useaan tyyppiin. Master datan tyyppejä ovat (McGilvray 2008, s. 42)

- tuotetiedot
- tilitiedot
- dokumentit
- työntekijätiedot
- asiakastiedot
- toimittajatiedot
- valmistajatiedot
- sijaintitiedot
- myyntialueet
- toimipistetiedot

Master datalle tyypillisiä piirteitä ovat stabiilius, kompleksisuus, uudelleenkäyttö, korkea arvo, elinkaari, riippumattomuus sekä läheisyys liiketoimintaan (Vilminko-Heikkinen & Pekkola 2012). Nämä ominaispiirteet ja niiden kuvaukset on esitetty taulukossa 2.1.

**Taulukko 2.1.** Master datan ominaispiirteet. (Vilminko-Heikkinen & Pekkola 2012)

| Ominaispiirre             | Kuvaus  |
|---------------------------|---|
| Stabiilius                | Master data muuttuu harvoin.  |
| Kompleksisuus             | Master data on monimuotoista ja sama data esiintyy organisaation eri toiminnoissa.            |
| Uudelleenkäyttö           | Master dataa käytetään tyypillisesti pitkään.   |
| Korkea arvo               | Master data on organisaation liiketoiminnan kannalta tärkeintä dataa, esimerkiksi tuotetieto. |
| Elinkaari                 | Master dataa luodaan, käytetään, muokataan ja poistetaan hallitusti.                          |
| Riippumattomuus           | Master data ei tarvitse muuta dataa ollakseen arvokasta.                                      |
| Läheisyys liiketoimintaan | Master data esiintyy tiiviisti transaktioiden yhteydessä.                                     |

Master datalla on siis varsin monipuoliset ominaispiirteet. Toisin kuin muut datakategoriat, master data on arvokasta itsessään ilman mitään muuta dataa. Se esiintyy myös todella suuressa roolissa yrityksen liiketoiminnassa, sillä jokainen transaktio sisältää master dataa, esimerkiksi myytävän tai ostettavan tuotteen sekä asiakkaan ja toimittajan tiedot.

Tässä työssä keskitytään tarkastelemaan master datan laatua ja siihen vaikuttavia tekijöitä. Datakategorioissa on hyvin paljon riippuvuuksia, kuten kuva 2.4 osoitti. Näin ollen datan laadun tarkastelu laajenee luontaisesti hieman myös muihin datakategorioihin. On kuitenkin huomattava, että kullakin datakategorialla on oma roolinsa organisaatiossa, minkä vuoksi kunkin hallinta tulisi toteuttaa tämä rooli huomioiden (Panian 2010). Tarkastelun pääpaino tässä työssä pyritään kuitenkin säilyttämään master datassa ja vielä tarkemmin tuotetiedossa ja dokumenteissa. Master datan tyypeistä tuotetieto on yksi haastavimmista tarkasteltavista datan laadun suhteen, sillä se on usein hyvin laaja-alaista ja sisältää muihin tyyppeihin verrattuna erityisen paljon attribuutteja (Silvola 2018, s39). Tuotetiedon katsotaankin olevan toiseksi alttein (asiakastiedon jälkeen) datan laadullisille ongelmille (Russom 2006). Tuotetietoa on myös luonteva tarkastella kohdeyrityksen kontekstissa, jossa tuotetieto on juuri siirretty uuteen PLM-järjestelmään.

### 3. DATAN LAADUN MÄÄRITELMÄ JA VIITEKEHYS

Laatua on yleisesti vaikea määritellä. Tämä pätee myös datan laadun suhteen. Yleisesti datan laatu määritellään sopivuutena tiettyyn käyttötarkoitukseen (engl. *fitness for use*) (Tayi & Ballou 1998). Datan laatu näyttäytyy eri tavoin eri ihmisille riippuen esimerkiksi käyttötarkoituksesta ja tilanteesta (Wand & Wang 1996). Tässä luvussa tarkastellaan kirjallisuudessa esiintyviä datan laadun osatekijöitä sekä rakennetaan kirjallisuuden pohjalta tähän työhön sopiva datan laadun viitekehys.

#### 3.1 Datan laadun osatekijät kirjallisuudessa

Kirjallisuudessa datan laadulle on listattu useita erilaisia osatekijöitä, joita ymmärtämättä ei voi käsitellä datan laadun ongelmia (Tayi & Ballou 1998). Tutkimuksissa käytettyjen osatekijöiden määrä vaihtelee todella suuresti muutamasta osatekijästä jopa yli kahteenkymmeneen. Taulukkoon 3.1. on kerätty lista erilaisista kirjallisuudessa esiintyvistä datan laadun osatekijöistä.

**Taulukko 3.1.** Datan laadun osatekijät. (koottu lähteistä Levitin & Redman 1995, Wand & Wang 1996, Strong et al. 1997, Tayi & Ballou 1998, Wang 1998, Lee et al. 2002, Pipino et al. 2002, Setia et al. 2013, Sebastian-Coleman 2013, Silvola 2018, Zhang et al. 2019)

|                    |                 |                 |                   |
|--------------------|-----------------|-----------------|-------------------|
| Ajankohtaisuus     | Arvokkuus       | Esitysmuoto     | Helppokäyttöisyys |
| Jatkuvuus          | Johdonmukaisuus | Joustavuus      | Kattavuus         |
| Kestävyys          | Käytettävyys    | Laajuus         | Maine             |
| Mielekkyys         | Objektiivisuus  | Oikea formaatti | Oikea määrä       |
| Rakenne            | Relevanttius    | Saatavuus       | Selkeys           |
| Tarkkuus           | Tehokkuus       | Tiiveys         | Tulkittavuus      |
| Tunnistettavuus    | Turvallisuus    | Tärkeys         | Uskottavuus       |
| Verrattavuus       | Virheettömyys   | Yhdenmukaisuus  | Yhtenäisyys       |
| Yksityiskohtaisuus | Ymmärrettävyys  |                 |                   |

Datan laadun ongelma on puute yhdessä tai useammassa näistä osatekijöistä, mikä jättää datan sopimattomaksi käyttötarkoitukseensa. Datan laadun ongelmien seurauksena data ei tuota lisäarvoa organisaatiolle. (Strong et al. 1997)

Osatekijöiden määrittelyä hankaloittaa osatekijöiden väliset riippuvuudet ja osittaiset päällekkäisyydet. Lisäksi datan laatuun liittyvä termistö ei ole täysin vakiintunutta, mikä vaikeuttaa kirjallisuuslähteiden vertailua. Datan laadun osatekijöiden perusrunko on pääsääntöisesti aina sama, mutta lähteiden välillä on huomattavia eroavaisuuksia (Wand & Wang 1996).

Useat datan laadun osatekijöitä listaavat lähteet pyrkivät myös luokittelemaan osatekijät. Myös luokitteluissa on paljon eroa lähteiden välillä. Silvola (2018) ja Wand & Wang (1996) jaottelevat tekijät melko karkeasti datan luonteenomaisiin tekijöihin sekä kontekstuaalisiin tekijöihin. Strong et al. (1997) ja Tayi & Ballou (1998) lisäävät tähän luokitteluun vielä esitysmuodolliset tekijät ja saatavuudelliset tekijät.

Eppler (2003) puolestaan jakaa datan laadun osatekijät informaatiotasoihin, jotka ovat infrastruktuuri, prosessi, tuote ja yhteisö. Infrastruktuuritasolla korostuvat tekniset ominaisuudet, kuten datan saatavuus ja turvallisuus. Prosessitasolla datan laatu muodostuu pääasiassa datan hyödyllisyydestä ja käytettävyydestä prosesseissa. Tuotetasolla keskiössä ovat tekijät, jotka kuvaavat datan sisältöä, kuten virheettömyys ja ajankohtaisuus. Yhteisötasolla tarkastellaan datan relevanttiutta ja ymmärrettävyyttä. Taulukkoon 3.2 on koottu karkea esitys erilaisista datan laadun osatekijöiden luokitteluista.

**Taulukko 3.2.** Datan laadun osatekijöiden luokittelu.

| Wand & Wang (1996), Silvola (2018) | Strong et al. (1997), Tayi & Ballou (1998) | Eppler (2003)   |
|------------------------------------|--|-----------------|
| Luonteenomaiset tekijät            | Luonteenomaiset tekijät                    | Tuote           |
| Kontekstuaaliset tekijät           | Kontekstuaaliset tekijät                   | Prosessi        |
|                                    |  | Yhteisö         |
|                                    | Esitysmuodolliset tekijät                  | Infrastruktuuri |
|                                    | Saatavuudelliset tekijät                   |                 |

Kuten taulukosta 3.2 huomataan, luokitteluissa on eroja, mutta jälleen perusrunko on sama. Varhaisissa datan laadun tutkimuksissa keskityttiin datan laadun sisäisiin osatekijöihin. Datan laadun tarkastelussa on kuitenkin tärkeää huomioida myös konteksti. (Foley & Helfert 2010) Kontekstuaalisia tekijöitä ovat esimerkiksi käyttäjät, tietojärjestelmät

sekä prosessit. Tärkein havainto taulukosta 3.2 on, että varsinainen data muodostaa hyvin pienen osan datan laadun kokonaisuudesta.

Datan laadun osatekijöiden luokittelulla saatava hyöty jää monessa tapauksessa melko vähäiseksi. Luokittelun tarve korostuu silloin, kun tarkasteltavana on erityisen suuri määrä luokiteltavia osatekijöitä. Tällöinkin luokittelu saattaa hankaloittaa tarkastelua, koska joitakin osatekijöitä voidaan tarkastella useammasta näkökulmasta. Tämä luo tarpeen luoda kaksi osatekijää, jotka tarkastelevat samaa asiaa kahdesta eri näkökulmasta.

### 3.2 Datan laadun viitekehys

Datan laadun osatekijöiden suuri määrä saattaa mahdollistaa laadullisen ongelman tar-  
kan kohdistamisen, mutta samalla se tekee datan laadun tarkastelusta työlästä ja han-  
kalaa. Lisäksi osatekijöiden suuren määrän käyttäminen edellyttää, että samankaltaisten  
tekijöiden väliset erot selitetään tarkasti. Tässä työssä pyritään lähtökohtaisesti intuitiivi-  
seen lähestymistapaan sekä minimoimaan osatekijöiden määrä yhdistelemällä tekijöitä  
asianmukaisesti. Tällöin myöskään varsinaiselle luokittelulle ei ole tarvetta.

Datan laadun osatekijöiden luokittelua voidaan kuitenkin tässä yhteydessä hyödyntää  
kääntämällä tarkastelu ympäri hyödyntämällä luokittelua kunkin käytetyn osatekijän nä-  
kökulmana. Nyt näkökulmiksi tulee valita sellaiset luokittelut, jotka eivät kuvaa suoraan  
mitään yksittäistä osatekijää (tällainen olisi esimerkiksi luokka ”saatavuudelliset tekijät”).

Taulukosta 3.2 havaitaan, että kaikissa esitetyissä luokitteluissa esiintyy datan luonteen-  
omaiset tekijät. Tämä on siis selkeä ensimmäinen näkökulma; kutsutaan sitä datanäkö-  
kulmaksi. Silvola (2018, s32) tunnistaa tutkimuksessaan master datan hallinnan kulma-  
kiviksi datan lisäksi tietojärjestelmät sekä liiketoimintaprosessit. Nämä toimivat hyvin da-  
tan laadun näkökulmina ja ne esiintyvät myös esimerkiksi Epplerin (2003) tutkimuk-  
sessa. Epplerin (2003) käyttämästä luokittelusta jäljelle jää vielä yhteisö eli organisaation  
henkilöstö, joka käyttää tarkasteltavaa dataa ja tietojärjestelmiä. Kutsutaan tätä käyttä-  
jänäkökulmaksi.

Datan laadun osatekijöiden näkökulmat ovat siis data, tietojärjestelmä, käyttäjä ja pro-  
sessi. Datanäkökulma kuvaa datan luonteenomaisia, sisäisiä tekijöitä, kuten tietomallia.  
Tietojärjestelmä kuvaa tekijöitä, jotka vaikuttavat datan laatuun tietojärjestelmän kautta,  
kuten esimerkiksi tietojärjestelmän hakuominaisuudet sekä joustavuus. Käyttäjiin liittyvät  
tekijät kuvaavat käyttäjien kokemaa datan laatua, kun taas prosessiin liittyvät tekijät ku-  
vaavat datan arvoa prosesseissa, joissa dataa käytetään. Neljä näkökulmaa voidaan ku-  
vata myös prosessimallina, kuten on tehty kuvassa 3.1.



**Kuva 3.1.** *Datan laadun näkökulmat prosessin vaiheina.*

Kuvasta 3.1 käy ilmi näkökulmien roolit datan elinkaareissa. Liiketoimintaprosessi luo tarpeen luoda dataa. Käyttäjä luo datan ja tallentaa sen tietojärjestelmään. Käyttötilanteessa data haetaan tietojärjestelmästä takaisin käyttäjälle, joka hyödyntää dataa liiketoimintaprosessissa. Kuva 3.1 on lähinnä suuntaa antava, sillä esimerkiksi jotkin prosessit tuottavat ja tallentavat datan automaattisesti ilman käyttäjää. Kuvan 3.1 keskeisin viesti on se, että datan laatu toimii kokonaisuutena. Jos ketjun yksi vaihe on heikkolaatuinen, se heijastuu suoraan myös muihin vaiheisiin.

Varsinaisten datan laadun osatekijöiden valitsemisessa voidaan lähteä alun perin jo vuonna 1985 tunnistetuista neljästä osatekijästä: tarkkuus, kattavuus, jatkuvuus ja ajantasaisuus (Ballou & Pazer 1985). Tarkkuuden ja jatkuvuuden voidaan katsoa toimivan hyvin valituissa näkökulmissa ja niiden kautta pystytään esittämään suuri osa taulukon 3.1 osatekijöistä. Sen sijaan kattavuus ei itsessään tuo lisäarvoa, jos tarkkuus ja saatavuus on huomioitu. Saatavuus puolestaan on taulukossa 3.2 nostettu jopa omaksi osatekijöiden luokakseen (saatavuudelliset tekijät), joten se toimii erinomaisesti tässä datan laadun osatekijänä. Kattavuus voidaan esittää saatavuuden datanäkökulmana. Vastavasti ajantasaisuus on osatekijänä melko kapea-alainen. Wand & Wang (1996) listaavat viideksi eniten siteeratuksi datan laadun osatekijäksi tarkkuuden, luotettavuuden, ajankohtaisuuden, relevanttiuden sekä kattavuuden. Näistä luotettavuuden voidaan katsoa toimivan hyvin valittujen näkökulmien kanssa. Relevanttius toimii tekijänä vain prosessin näkökulmasta. Tarkasteltaessa taulukkoa 3.1 näiden neljän osatekijän (tarkkuus, jatkuvuus, saatavuus ja luotettavuus) ja valittujen näkökulmien kautta, suurin osa taulukon osatekijöistä pystytään kartoittamaan. Jäljelle jäävät turvallisuuteen viittaavat tekijät, joten valitaan turvallisuus viidenneksi osatekijäksi.

Taulukossa 3.3 on esitetty viisi datan laadun osatekijää sekä kunkin neljä näkökulmaa. Taulukossa on hyödynnetty taulukon 3.1 osatekijöitä havainnollistamaan näkökulmien sisältöä.

**Taulukko 3.3.** *Datan laadun osatekijät vasemmalla ja niiden näkökulmat ylärivillä.*

|                     | Data   | Tietojärjestelmä   | Käyttäjä   | Prosessi  |
|---------------------|--|--|--|---|
| <b>Tarkkuus</b>     | Tietomalli, rakenne,<br>laajuus,<br>yksityiskohtaisuus,<br>tiivius | Joustavuus   | Tulkittavuus,<br>ymmärrettävyys                        | Relevanttius, tärkeys                             |
| <b>Saatavuus</b>    | Kattavuus  | Datan määrä,<br>hakuominaisuudet,<br>esitysmuoto             | Käytettävyys,<br>helppokäyttöisyys,<br>tunnistettavuus | Oikea data oikeassa<br>paikassa oikeaan<br>aikaan |
| <b>Jatkuvuus</b>    | Johdonmukaisuus,<br>yhdenmukaisuus                                 | Ei duplikaatteja,<br>integraatio,<br>verrattavuus            | Selkeys  | Arvokkuus,<br>tehokkuus                           |
| <b>Luotettavuus</b> | Virheettömyys,<br>objektiivisuus                                   | Seurattavuus,<br>historia                                    | Maine, uskottavuus                                     | Ajantasaisuus                                     |
| <b>Turvallisuus</b> | Kestävyys  | Pääsy järjestelmään,<br>käyttöoikeudet,<br>muutoksenhallinta | Omistajuus   | Käyttövarmuus                                     |

Tällä mallilla päästään hyvin hallittavaan, mutta samalla monipuoliseen, datan laadun kokonaiskuvaan. Mallilla myös vältetään tarve kunkin näkökulman omille osatekijöille, jotka usein termeinä ovat lähes synonyymejä, ja joiden tulkinta perustuu käytännössä pelkästään termien selityksiin.

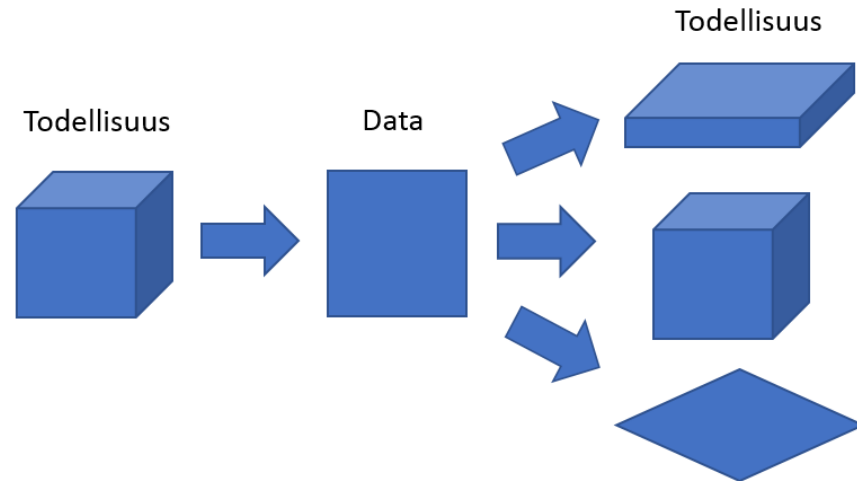
### 3.3 Datan laadun osatekijöiden kuvaukset

#### Tarkkuus

Tarkkuus tarkoittaa, että datalla pystytään kuvaamaan todellisuutta riittävällä tarkkuudella ja laajuudella (Wand & Wang 1996). Wand & Wang (1996) nostavat esille, että datalla pyritään aina kuvaamaan jotakin todellisuuden asiaa tai tapahtumaa. Näin ollen datan laatu on keskeisesti sidoksissa siihen, miten hyvin yhteys todellisuuden ja datan välillä toimii kumpaankin suuntaan. Tarkkuuden kannalta on tärkeää, että datan rakenne ja tietomalli on suunniteltu vastaamaan datan käytön tarpeita. Tämä on tarkkuuden datanäkökulma.

Käyttäjän näkökulma korostuu otettaessa dataa ulos tietojärjestelmästä tai tulkittaessa sitä. Todellisuus pitää pystyä mallintamaan datana kattavasti, mutta myös datan on kuvattava yksiselitteisesti tiettyä todellisuutta. Jos datan tarkkuus on käyttäjän näkökulmasta heikko, se ei kuvaa todellisuutta yksiselitteisesti, jolloin data on tulkinnanvaraista. (Wand & Wang 1996) Tätä on havainnollistettu kuvassa 3.2. Ongelma korostuu entistään, sillä yksi käyttäjä saattaa tulkita datan eri tavalla kuin toinen käyttäjä.





**Kuva 3.2.** *Epäonnistunut todellisuuden tallentaminen dataksi.*

Prosessin näkökulmasta datan tarkkuus tarkoittaa sitä, että datalla on kuvattu oikeita asioita. Vaikka data kuvaisi todellisuutta täydellisesti ja ilman mitään tulkinnanvaraa, datan arvo liiketoiminnan kannalta on hyvin pieni, jos kuvauksen kohde on väärä. Uusien tarpeiden ilmaantuessa tietojärjestelmän on oltava joustava ja mahdollistettava datan uusien tarpeiden tallentaminen sekä esittäminen.

### Saatavuus

Saatavuus on yksi tärkeimmistä datan laadun kontekstisidonnaisista osatekijöistä ja se vaikuttaa useisiin muihin tekijöihin. Kontekstisidonnaisuus tarkoittaa sitä, että saatavuutta on aina tarkasteltava jossakin tietyssä kontekstissa. (Foley & Helfert 2010) Datan saatavuus kuvaa, miten helposti dataan pääsee käsiksi ja miten nopea prosessi datan hakeminen on. Datanäkökulmasta hyvä saatavuus edellyttää datan kattavuutta. Tämä tarkoittaa, että datan tietorakenne on täytetty kunnolla, sillä tyhjät tietokentät eivät mahdollista tehokasta hakutulosten suodatusta.

Tietojärjestelmänäkökulmasta saatavuutta kuvaa tietojärjestelmän mahdollistamat haku-toiminnot. Tietyn datan löytäminen on aina hankalampaa, jos datan haku on suoritettava yhdellä tietyllä tavalla, esimerkiksi hakemalla käyttäen yksilöivää tunnistetta. Lisäksi saatavuuteen kuuluvat myös datan asianmukainen esitystapa ja tietojärjestelmän käyttäjä-rajapinnan ulkoasu (Foley & Helfert 2010), jotka parantavat datan tunnistettavuutta. Jos dataa pystytään tietojärjestelmän kautta tarkastelemaan esimerkiksi graafisessa tai visuaalisessa muodossa, jotka tukevat toisiaan, datan käytettävyyks paranee.

Saatavuuden tietojärjestelmänäkökulmaan liittyy oleellisesti myös tietojärjestelmässä olevan datan määrä. On ilmiselvää, että datan liian vähäinen määrä aiheuttaa ongelmia,

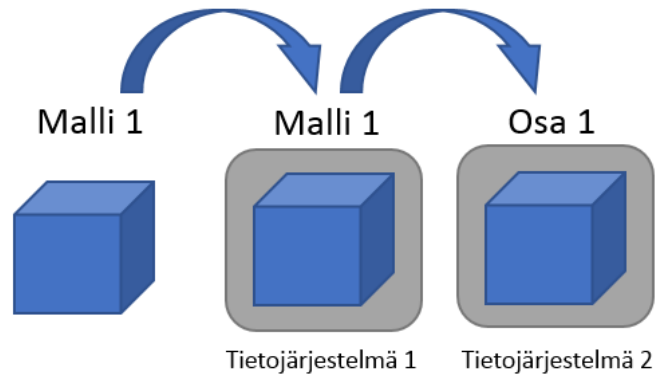
mutta ongelmaksi muodostuu myös datan liian suuri määrä. Tarpeeton data on todella yleistä yrityksille (Redman 1998). Suuret datamäärät hankaloittavat helposti datan saatavuutta, haudaten oikean datan ylimääräisen datan joukkoon. Silver (2012, s152) huomauttaa, että tällaisissa järjestelmissä käyttäjät usein myös tekevät väärä johtopäätöksiä pohjautuen ylimääräiseen dataan, erityisesti dataa analysoidessa. Yksi datan laadun tutkimuksen keskeisistä tavoitteista onkin minimoida tarpeettoman datan vaikutuksia (Foley & Helfert 2010). Tarpeettoman datan poistaminen vähentäisi tietojärjestelmän kuormaa, helpottaisi tietojärjestelmän ylläpidon työtä ja vähentäisi riskiä, että data päätyisi ulkopuolisille tahoille.

Perinteiset lähestymistavat käsittävät datan saatavuuden ongelmat usein teknisiksi ongelmiksi laadullisten ongelmien sijaan. Tällöin huomio kiinnittyy muun muassa tietojärjestelmän toimintaan ja järjestelmään pääsyyn. Datan käyttäjien näkökulmasta kuitenkin saatavuuteen kuuluu oleellisesti käytön ja datan hakemisen helppous. (Strong et al. 1997) Vaikka saatavuus painottuu kontekstinäkökulmiin (tietojärjestelmä ja käyttäjät) varsinaisen datan sijaan, useat tutkimukset ovat osoittaneet sillä olevan merkittävää vaikutusta datan laatuun.

Prosessinäkökulmasta saatavuus voidaan nähdä osittain muiden näkökulmien summana. Tällöin saatavuus tarkoittaa sitä, että oikea data pystytään löytämään ja ottamaan käyttöön prosessissa, kun sitä tarvitaan, hyödyntäen käytettävissä olevaa tietojärjestelmää sekä henkilöstöä.

### **Jatkuvuus**

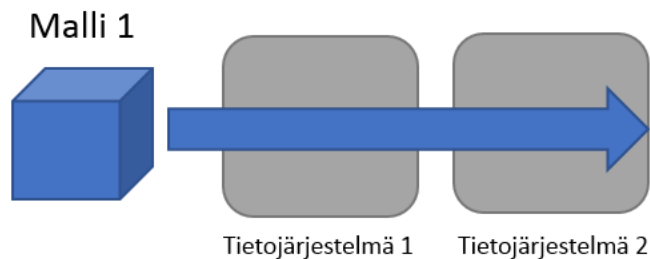
Jatkuvuus tarkoittaa sitä, että kukin datanimike määritellään ja hallitaan vain yhdessä paikassa, eikä samaa asiaa kuvaavissa datanimikkeissä ole keskenään eroa. Datanäkökulmasta tämä tarkoittaa johdonmukaisuutta, joka merkitsee, että esimerkiksi datan kuvauksissa käytettävät merkinnät noudattavat samaa kaavaa ja kieli on kaikkialla sama. Samaa tarkoittavat todelliset arvot kuvataan tietomallissa aina samoilla datan arvoilla (Wand & Wang 1996). Kuvassa 3.3 on esitetty epäjohdonmukainen tilanne, jossa aluksi on luotu ”Malli 1”, joka on viety kahteen tietojärjestelmään. Malli on kaikissa sama, mutta jälkimmäisessä tietojärjestelmässä malli tunnetaan nimellä ”Osa 1”.



**Kuva 3.3.** Epäjohdonmukainen nimeäminen.

Kuvasta 3.3 havaitaan nopeasti myös toinen jatkuvuuden ongelma. Kuvassa sama data esiintyy omana objektinaan kolmessa sijainnissa. Tällaisissa tilanteissa jatkuvuuteen liittyy myös datan virheettömyys ja ajantasaisuus, sillä data voi olla oikein ja ajan tasalla yhdessä järjestelmässä, mutta virheellistä ja vanhentunutta toisessa järjestelmässä. Tällöin data on siis päivitettävä erikseen kahteen eri paikkaan.

Tietojärjestelmänäkökulmasta jatkuvuus tarkoittaa, että datalla ei ole epäjatkuvuuskohtia esimerkiksi eri prosessien tai tietojärjestelmien välillä. Organisaatioilla on suuri tarve integroida erilaiset tietojärjestelmänsä aikaisempaa paremmin (Martio 2015, s9). Silvola (2018) mainitsee, että parhaimmillaan data määritellään vain kerran, minkä jälkeen sitä käytetään muuttumattomana eri prosesseissa. Tätä on havainnollistettu kuvassa 3.4, jossa kuvan 3.3 tilanne on korjattu integroimalla tietojärjestelmät 1 ja 2 käyttämään alkuperäistä mallia.



**Kuva 3.4.** Data määritellään kerran ja sitä käytetään muuttumattomana useissa järjestelmissä.

Jatkuvuutta tarkastellessa on tärkeää huomata, että kyseessä ei ole välttämättä ainoastaan yrityksen omat, sisäiset tietojärjestelmät. Yrityksillä on paljon yhteistyökumppaneita, joille yritys toimittaa dataa. Jos dataa toimittavan yrityksen ja dataa vastaanottavan yhteistyökumppanin välillä vallitsee eriävät näkemykset datan laadun suhteen, yhteistyö-

kumppani saattaa alkaa pitää datasta omaa tietokantaa. Tällöin näiden yritysten tietojärjestelmien välille muodostuu datan epäjatkuvuuskohta. Vastaavia epäjatkuvuuksia voi muodostua myös saman yrityksen eri toimipisteiden välille. (Redman 1998)

Jatkuvuuden tietojärjestelmänäkökulmassa on huomioitava myös duplikaattinimikkeiden merkitys. Duplikaattinimikkeet heikentävät datan laatua monella tavalla. Esimerkiksi käyttäjät eivät välttämättä tiedä, mitä duplikaateista pitäisi käyttää ja mikä duplikaateista on ajan tasalla. Lisäksi tilauksia ja varastoa tarkasteltaessa on vaarana, että väärä duplikaattinimike ilmoittaa varastomäärän olevan liian vähäinen, vaikka oikeaa nimikettä on varastossa riittävästi. Tällöin nimikettä tilataan duplikaatin pohjalta lisää ja varastoon päätyykin liian suuri määrä nimikettä.

Käyttäjän näkökulmasta datan jatkuvuus merkitsee selkeyttä. Epäjohdonmukaiset merkinnät aiheuttavat sekaannusta ja tulkinnanvaraisuutta.

Datan jatkuvuus ei ehkä ole tärkein datan laadun osatekijä, olettaen, että muut osatekijät on asianmukaisesti huomioitu, mutta se nousee kuitenkin liiketoiminnan kannalta tärkeäksi tekijäksi. Datan epäjatkuvuus aiheuttaa suoraan lisäkustannuksia mm. datan ylläpidossa sekä hidastaa liiketoiminnan muutoksiin reagoimista (Silvola 2018). Datan jatkuvuudella on myös vaikutuksia muihin datan laadun osatekijöihin. Prosessinäkökulmasta jatkuvuus onkin koko tietovirran ja tietojärjestelmäkokonaisuuden kustannustehokkuutta.

### **Luotettavuus**

Datan luotettavuuteen kuuluu, että data on virheetöntä ja objektiivista. Siinä ei esiinny vinoumia (Wand & Wang 1996) tai subjektiivisia näkemyksiä. Varsinaiset virheet (mukaan lukien epä johdonmukaisuudet) ovat hyvin yleinen datan laadun ongelma. Ne johtavat moniin liiketoiminnan ongelmiin, kuten laskutuksen ja toimitusten virheisiin, menetettyihin liiketoimintamahdollisuuksiin sekä virheellisiin päätöksiin. (Silvola 2018) Redman (1998) esittää, että tyypillisen yrityksen datan virheellisten tietokenttien määrä kaikista tietokentistä on keskimäärin 1-5 %.

Tietojärjestelmäkontekstissa sen sijaan datan luotettavuuteen liittyy oleellisesti todennäköisyys tulevien virheiden välttämiseen (Wand & Wang 1996). Hyvin suunniteltu tietojärjestelmä pystyy siis ohjaamaan dataa laadukkaampaan suuntaan. Nykytilaa kuvaavan datan lisäksi laadukkaaseen dataan kuuluu myös historiatieto, joka mahdollistaa seurattavuuden. Historiatietojen perusteella voidaan tunnistaa datassa käytetyt uudet tietolähteet sekä päivitykset (Silvola 2018). Tämä johtaa siihen, että dataan kohdistuvat toimenpiteet sekä muutokset ovat aina jäljitettävissä.

Hallittu muutoksenhallinta mahdollistaa osaltaan myös sen, että dataan ei pääse virheitä ja epäjohtonmukaisuuksia. Laadukas muutoksenhallintaprosessi tuottaa laadukasta dataa.

Käyttäjänäkökulmasta luotettavuuteen liittyvät datan maine sekä uskottavuus. Maine kuvaa organisaatiossa yleisesti vallitsevaa näkemystä datasta ja sen laadusta, kun taas uskottavuus on yksilökohtainen subjektiivinen tekijä. Se kuvaa, miten yksilö suhtautuu dataan. Strong et al. (1997) mainitsevatkin, että datan laatua ei voi tarkastella täysin dataa käyttävistä yksilöistä erillisenä kokonaisuutena.

Luotettavuus pitää myös sisällään datan lähteen luotettavuuden. Vaikka data itsessään olisi laadukasta, heikkolaatuinen lähde saattaa vähentää datan koettua laatua käyttäjän silmissä. Vastaavasti organisaation historia datan laadun suhteen saattaa heijastua datan luotettavuuteen. Aiemmin koetut ongelmat saattavat vähentää datan koettua laatua tulevaisuudessa, vaikka kyseessä ei välttämättä olisi enää edes sama data.

Prosessin näkökulmasta datan virheettömyys kantaa jo pitkälle, mutta sen lisäksi datan on myös oltava ajan tasalla. Ajantasaisuus kuvaa sitä, kuinka hyvin data kuvaa sen hetkistä todellisuutta. Laadukkaan datan on pystyttävä kuvaamaan todellisuutta kullakin hetkellä muuttuvassa ympäristössä (Wand & Wang 1996, Silvola 2018).

## **Turvallisuus**

Turvallisuus kuvaa datan alttiutta epäsuotuisille muutoksille tai väärille käyttäjille. Datanäkökulmasta tätä voidaan kuvailla datan kestävyudeksi. Kestävyys tarkoittaa, miten helposti dataan pääsee virheitä ja miten helposti data voi hajota. Kestävyyttä voidaan parantaa tietomallin avulla esimerkiksi määrittelemällä valmiit lista-arvot mahdollisimman moniin tietokenttiin. Tällöin tietomalli ohjaa käyttäjän täyttämään kentän oikein. Toisaalta kestävyteen kuuluu myös se, kuinka helposti datanimikkeen voidaan katsoa olevan virheellinen.

Turvallisuuden tietojärjestelmänäkökulma pitää sisällään käytön ja saatavuuden rajoittamisen vain tietyille käyttäjille, esimerkiksi organisaation työntekijöille. Verkossa käsiteltävän datan saatavuuden rajoittaminen on todella tärkeää (Foley & Helfert 2010). Datan käytön (lisääminen, muokkaaminen ja poistaminen) hallinta on yksi turvallisuuden kulmakivistä, jotta esimerkiksi käyttäjät eivät vahingossa pääsisi poistamaan dataa. Sen lisäksi, että ulkopuolisella ei saa olla pääsyä dataan, on myös tärkeää, että työntekijöillä ei ole pääsyä tai käyttöoikeuksia heille epärelevanttiin dataan. Jos käyttäjillä on liikaa oikeuksia, kasvaa todennäköisyys vahingossa tehtyihin epäsuotuisiin muutoksiin.

Käyttäjänäkökulmasta datalla on oltava selkeästi määritelty omistaja, joka on datasta vastuussa. Datan omistajuuteen kuuluu kirjallisuudessa useita rooleja, mutta tämän työn kontekstissa riittää, että datalle on määritelty omistaja.

Prosessinäkökulmasta turvallisuudessa korostuu käyttövarmuus. Kaiken datan käsittely keskitetysti yhdessä tietojärjestelmässä asettaa organisaatiolle järjestelmäriskin. Jos tietojärjestelmä pettää, organisaatio voi pahimmillaan jäädä täysin toimintakyvyttömäksi.

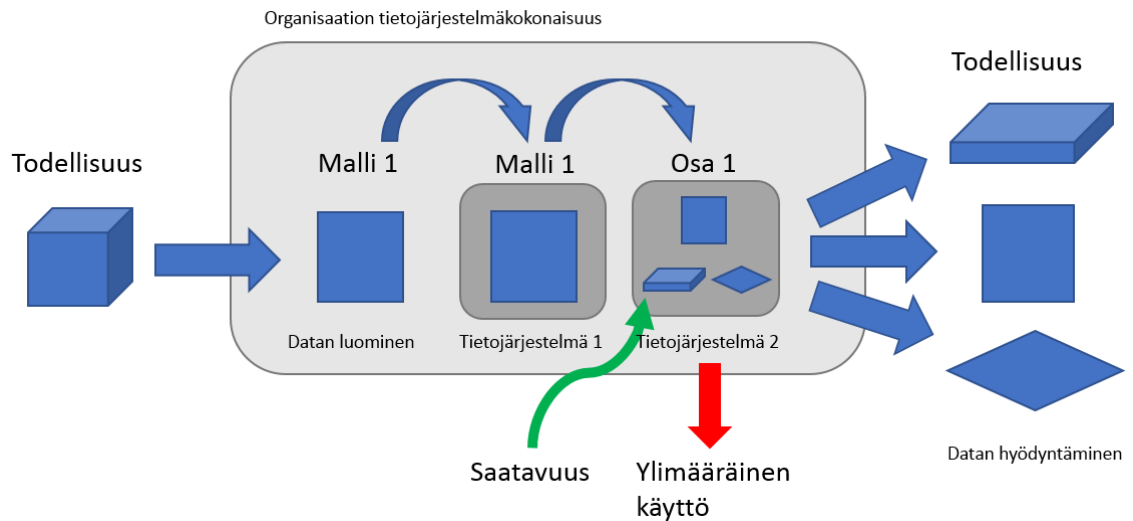
Kuten aiemmin on mainittu, datan laadun osatekijöiden määrittelyssä on hyvin paljon eroa eri kirjallisuuslähteissä. Yllä esitetty jako vain viiteen laadun osatekijään ja niiden neljään näkökulmaan luo helposti lähestyttävän ja ymmärrettävän viitekehyksen datan laadun muodostumisesta. On kuitenkin huomattava, että taulukossa 3.3 esitetyt kuvaukset eivät ole kaiken kattavia. Lisäksi useissa viitekehyksen kohdissa on selvää päällekkäisyyttä. Taulukkoon 3.4 on vielä koottu datan laadun viisi osatekijää sekä niiden tiivistetyt kuvaukset.

***Taulukko 3.4. Datan laadun osatekijät.***

| <b>Datan laadun osatekijä</b> | <b>Selitys</b>   |
|-------------------------------|--|
| Tarkkuus                      | Datalla pystytään tiettyssä tietojärjestelmäkontekstissa kuvaamaan ymmärrettävästi relevanttia todellisuutta riittävällä tarkkuudella ja laajuudella.  |
| Saatavuus                     | Data on tietomallissaan kattavaa ja näin ollen helposti haettavissa ja käytettävissä tietojärjestelmästä.  |
| Jatkuvuus                     | Data ei sisällä epä johdonmukaisuuksia, eikä sillä ole epä jatkuvuuskohtia eri prosessien ja järjestelmien välillä, eikä data näin ollen aiheuta prosessille ylimääräisiä kustannuksia tai ongelmia. |
| Luotettavuus                  | Käyttäjä voi luottaa tietojärjestelmästä saadun datan olevan käyttökelpoista ja paikkansa pitävää.   |
| Turvallisuus                  | Kerran laadukkaaksi todettu data ei herkästi muutu heikkolaatuiseksi.  |

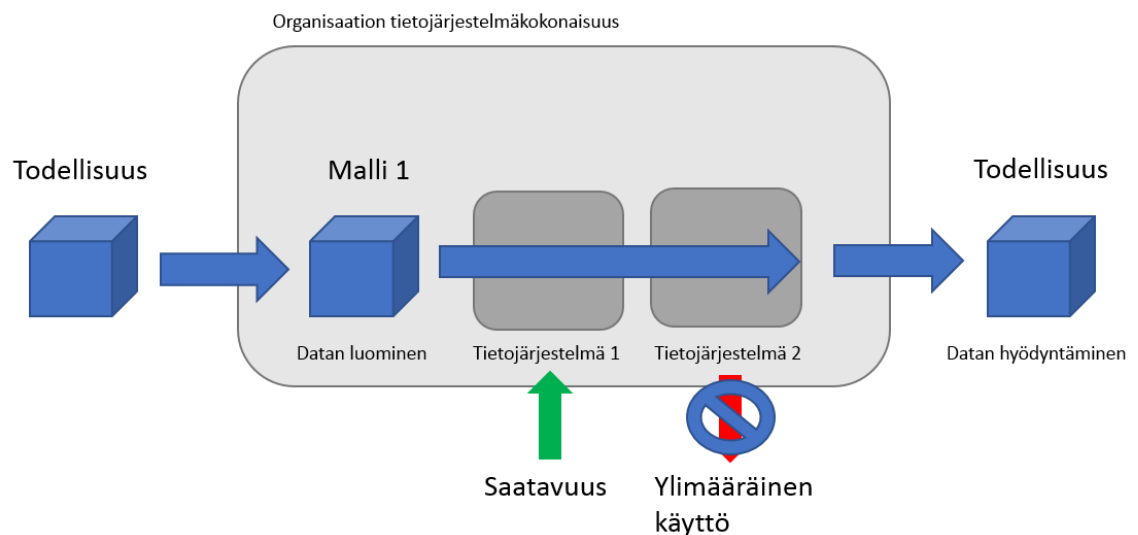
Kuvissa 3.5 ja 3.6 on vielä tiivistetty datan laadun keskeiset tekijät kahteen skenaarioon. Kuvassa 3.5 esitetään erityisen huono skenaario, jossa luotu data ei vastaa todellisuutta,

data ei ole johdonmukaista eikä jatkuvaa, järjestelmästä löytyy tarpeetonta dataa, tiedonhaku ei ole suoraviivaista, järjestelmä sallii ylimääräisen käytön (esim. ulkopuolisilla on pääsy järjestelmään) ja järjestelmästä saatava data on tulkinnanvaraista.



**Kuva 3.5. Huono skenaario.**

Kuvassa 3.6 puolestaan on havainnollistettu sama skenaario, kun datan laatu on korkea. Luotu data vastaa hyvin todellisuutta, data määritellään kerran ja sitä käytetään muuttumattomana läpi järjestelmän, tiedonhaku on suoraviivaista ja ylimääräinen käyttö on esitetty sekä järjestelmästä saatava data on yksiselitteistä.



**Kuva 3.6. Hyvä skenaario.**

Kuten kuvista havaitaan, datan laadulliset ongelmat ovat monitahoisia. On huomattava, että kuvissa 3.5 ja 3.6 ei oteta kaikkia datan laadun osatekijöitä huomioon. Kuvat eivät ota kantaa esimerkiksi ajantasaisuuteen ja uskottavuuteen. Toinen kuvista välittyvä ajatus on, miten huomattavasti datan jatkuvuus vaikuttaa muihin laadun tekijöihin.

Luvun 3 tiivistelmänä voidaan esittää taulukko 3.3 siistittynä siten, että kunkin laadun osatekijän näkökulmaan on valittu vain parhaiten sitä esittävä/esittävät kuvaukset. Lisäksi lisätään taulukkoon osatekijöiden yleiskuvaukset taulukosta 3.4. Tiivistelmätaulukko on esitetty taulukkona 3.5.

**Taulukko 3.5. Datan laadun viitekehys.**

|  | Data                          | Tietojärjestelmä                             | Käyttäjä           | Prosessi                                    |
|--|-------------------------------|--|--------------------|---|
| <b>Tarkkuus</b><br>Datalla pystytään tietyssä tietojärjestelmä kontekstissa kuvaamaan ymmärrettävästi relevanttia todellisuutta riittävällä tarkkuudella ja laajuudella.   | Tietomallin toimivuus         | Joustavuus                                   | Tulkittavuus       | Relevanttius                                |
| <b>Saatavuus</b><br>Data on tietomallissaan kattavaa ja näin ollen helposti haettavissa ja käytettävissä tietojärjestelmästä.  | Kattavuus                     | Haun toimivuus ja datan määrä                | Helppokäyttöisyys  | Oikea data oikeassa paikassa oikeaan aikaan |
| <b>Jatkuvuus</b><br>Data ei sisällä epäjohdonmukaisuuksia, eikä sillä ole epäjatkuvuuskohtia eri prosessien ja järjestelmien välillä, eikä data näin ollen aiheuta prosessille ylimääräisiä kustannuksia tai ongelmia. | Johdonmukaisuus               | Integraation toimivuus ja duplikaattittomuus | Selkeys            | (Kustannus)tehokkuus                        |
| <b>Luotettavuus</b><br>Käyttäjä voi luottaa tietojärjestelmästä saadun datan olevan käyttökelpoista ja paikkansa pitävää.  | Virheettömyys, objektiivisuus | Seurattavuus                                 | Maine, uskottavuus | Ajantasaisuus                               |
| <b>Turvallisuus</b><br>Kerran laadukkaaksi todettu data ei herkästi muutu heikkolaatuiseksi.   | Kestävyys                     | Käyttöoikeudet ja muutoksenhallinta          | Omistajuus         | Käyttövarmuus                               |



## 4. MASTER DATAN LAADUN HALLINTA

Master datan hallinta (*engl. Master Data Management, MDM*) on toimintojen kokonaisuus, jonka tavoitteena on luoda tarkka, ajankohtainen, kattava ja integroitu datavaraisto organisaation hallinnan ja kasvun tueksi (Berson & Dubov, 2007) sekä ylläpitää tätä datavaraistoa (Martio 2015, s95). Master datan hallinnassa kehitetään työkaluja, joilla eri tietokannoissa ja tietojärjestelmissä olevat nimiketiedot voidaan yhtenäistää (Martio 2015, s95).

Master datan hallintaan kuuluu datan laadun hallinta, omistajuuden määrittely sekä muutoksenhallinta (Silvola 2018, s41). Master datan laadun hallintaan on kehitetty useita prosessimalleja ja viitekehyskiä. Tällaisen prosessin tavoite on maksimoida datan laatu ja minimoida huonosta datasta aiheutuvat negatiiviset vaikutukset (Maydanchik 2007, s23). Martio (2015, s95) tuo esiin, että erityisesti suurissa globaaleissa organisaatioissa tehottomuus materiaaalivirtojen logistiikassa liittyy oleellisesti nimikkeiden hallintaan eli master dataan. Tässä luvussa kuvataan, millainen datan laadunhallintaprosessi on ja miten edellisen luvun datan laadun viitekehystä voidaan hyödyntää tässä prosessissa.

### 4.1 Datan laadunhallintaprosessimalli

Vaikka datan laadun hallinta- ja kehittämisprosesseja on useita, ne kuitenkin rakentuvat pääosin samoista vaiheista. Vaiheet ovat karkeasti taustan kartoitus, toimenpidesuunnitelma, suunnitelman toteutus ja monitorointi.

Ensimmäinen vaihe alkaa liiketoiminnan tavoitteiden määrittelyllä, jossa selvitetään mitä datalla halutaan saavuttaa ja mitä dataa on syytä tarkastella (McGilvray 2008). Tämän jälkeen kerätään tarvittava data ja arvioidaan sen laatu (Maydanchik 2007, s23; McGilvray 2008). Sekä McGilvray (2008) että Maydanchik (2007, s23) tuovat esiin, että tässä yhteydessä on syytä myös arvioida, mikä vaikutus tarkasteltavalla datalla ja sen laadulla on liiketoimintaan. Lisäksi Maydanchik (2007, s27) huomauttaa, että datan laadun arvioiminen manuaalisesti ei useimmissa tapauksissa ole toteutettavissa järkevällä työmäärällä. Tällöin muodostetaan datan laatusääntöjä, joiden avulla datan laatu voidaan kartoittaa automaattisesti. Säännöt pitävät sisällään esimerkiksi tietokentässä hyväksyttävät arvot. Kerran muodostettuja sääntöjä pystytään käyttämään yhä uudelleen, mutta on huomattava, että myös niiden laatu heikkenee ajan myötä ja niitäkin on ylläpidettävä (Maydanchik 2007, s28).

Prosessin toinen vaihe on toimenpiteiden suunnittelu. Vaihe alkaa kartoittamalla aiemmassa vaiheessa havaittujen datan laadun ongelmien taustasyitä. Tämän jälkeen muodostetaan suunnitelma havaittujen ongelmien korjaamiseksi ja taustasyiden poistamiseksi. (McGilvray 2008) Taustasyiden poistaminen on usein todella hankalaa, joten Maydanchik (2007, s24) kehottaa tässä vaiheessa keskittymään erityisesti datan epäjatkuvuuskohtiin, eli integraation rajapintoihin, konversioihin sekä datalähteiden yhdistymiin, sillä nämä tarjoavat parhaan hyödyn suhteessa käytettyyn työhön.

Kolmannessa vaiheessa toteutetaan toimenpidesuunnitelma. Vaihe pitää sisällään datan ongelmien korjauksen ja syiden poistamisen. Poistamalla datan ongelmia aiheuttavat syyt ennaltaehkäistään tulevia datan laadun ongelmia (Maydanchik 2007, s23; McGilvray 2008). Olemassa olevan datan korjaamisessa ei usein ole järkevää pyrkiä täydellisyyteen, vaan tavoite on optimoituva saatavan hyödyn ja käytetyt työn suhteen (Maydanchik 2007, s24).

Viimeinen vaihe on monitorointi, jossa tarkastellaan toteutettujen toimenpiteiden toimivuutta. Viimeiseen vaiheeseen kuuluu myös toimenpiteiden dokumentointi ja viestintä (McGilvray 2008), tosin McGilvray (2008) korostaa näiden tärkeyttä myös koko prosessin läpi. Useimmissa tilanteissa datan laadulla on tapana heikentyä, vaikka dataa ei edes muokattaisi. Tätä kutsutaan datan rappeutumiseksi. (Maydanchik 2007, s22) Jatkuva monitorointi on siis erittäin tärkeää uusien ongelmien havaitsemiseksi ja tarvittaessa datan laatu prosessin uudelleenkäynnistämiseksi (Maydanchik 2007, s23).

Tätä yleistä prosessimallia voidaan soveltaa luvussa 3 rakennetun datan laadun viitekehysten kanssa. Koska datan laadun ongelmien todettiin olevan puutteita yhdessä tai useammassa datan laadun osatekijässä, käy järkeen keskittää prosessi osatekijöiden ympärille. Laadunhallintaprosessin toteuttaminen kerralla koko viitekehykselle olisi suuri ja raskas operaatio. Tässä pystytäänkin hyödyntämään osatekijöiden näkökulmia. Valitsemalla yksi neljästä näkökulmasta kerrallaan tarkasteltavaksi, tarkastelu rajataan selkeään kokonaisuuteen, jota on helpompi käsitellä kuin esimerkiksi yhtä osatekijää kaikista näkökulmista.

Erityisesti tietojärjestelmien käyttöönotossa ja datamigraatiossa on syytä kiinnittää huomiota datan laadun neljään näkökulmaan. Datanäkökulmassa tunnistetut ongelmat on syytä korjata ennen datamigraatiota, muuten ongelmat vain siirtyvät uuteen järjestelmään. Tietojärjestelmän näkökulman ongelmat ovat todennäköisesti yksi keskeinen syy uuden tietojärjestelmän käyttöönotolle. Onkin siis tärkeä huomioida, että tietojärjestelmän näkökulman ongelmat saadaan korjattua uudella järjestelmällä. Käyttäjän näkökulman

ongelmia on hankalampi käsitellä tietojärjestelmä uudistuksen yhteydessä, koska käyttäjien ongelmat hyvin todennäköisesti muuttuvat siirryttäessä uuteen järjestelmään. Ongelmien esiintymistä voi kuitenkin pyrkiä ennakoimaan ja ennaltaehkäisemään koulutuksella, tiedottamisella ja käyttäjien motivoinnilla. Prosessinäkökulman ongelmien korjaaminen edellyttää, että datan käsittelyprosessi (datan luominen, käyttäminen ja hallinnointi) on määritelty ja sidottu liiketoimintaprosessiin. Sekä datan, tietojärjestelmien että käyttäjien on pystyttävä toimimaan laadukkaasti prosessin näkökulmasta tietojärjestelmä uudistuksen jälkeen. Taulukkoon 4.1 on koottu vielä datan laadun näkökulmat sekä kunkin näkökulman ominaispiirteet ja toimenpiteet tietojärjestelmä uudistuksessa.

**Taulukko 4.1.** *Datan laadun näkökulmien ominaispiirteet tietojärjestelmä uudistuksessa sekä näkökulmien toimenpiteet.*

|                        | Data   | Tietojärjestelmä   | Käyttäjä  | Prosessi  |
|------------------------|--|--|---|---|
| <b>Ominaispiirteet</b> | Korjaamattomat ongelmat siirtyvät uuteen järjestelmään | Keskeinen syy tietojärjestelmä uudistukselle                             | Uusi järjestelmä tuo mukanaan uudet ongelmat  | Asettaa raamit muille datan laadun näkökulmille   |
| <b>Toimenpiteet</b>    | Data on valmisteltava huolella ennen migraatiota       | Uusi järjestelmä on valittava ja konfiguroitava korjaamaan nämä ongelmat | Ongelmien ennakoiminen ja ennaltaehkäisy koulutuksen, tiedottomisen ja motivoinnin avulla | Prosessin on oltava selkeästi määritelty, jotta muut näkökulmat täyttävät prosessin vaatimukset |

Prosessimallin ensimmäisessä vaiheessa valitaan siis tarkasteltava näkökulma ja käydään läpi, missä datan laadun osatekijöistä on havaittu ongelmia. Kuhunkin näkökulmaan, jotka jäsennettiin luvussa 3 useasta eri lähteestä, liittyy omat keskeiset ominaispiirteensä. Nämä ominaispiirteet ohjaavat taulukon 4.1 mukaisesti tietojärjestelmä uudistuksessa tietyn tyyppisiin toimenpiteisiin. Seuraavaksi käydään läpi kunkin näkökulman keskeiset ominaispiirteet laadunhallintaprosessin kannalta.

## 4.2 Datan laadun näkökulmien ominaispiirteet

### Data

Datanäkökulmaan liittyvät sisäiset laadun osatekijät lähtevät oletuksesta, että datalla voidaan katsoa olevan laatuaspekti ilman kontekstia (Wang 1998). Loshin (2011) toteaa, että nämä tekijät vaikuttavat datan sisäiseen arvoon.

Datan tarkkuuteen vaikuttaa datassa käytetty tietomalli, jonka laadukkuutta on hankala arvioida. Tietomallin laadukkuus selviää parhaiten haastattelujen ja käyttäjä- ja prosessinäkökulmien kautta. Toisaalta tietomallista voidaan tarkastella tietokenttiä ja niiden määrää. Martio (2015, s95-96) listaa, että jokaisella nimikkeellä tulee olla yksi ja vain yksi päätunniste. Lisäksi nimikkeeltä on löydettävä havainnollistava kuvaus, tarkentava spesifikaatio ja nimikkeen on ilmennettävä jotakin määriteltyä tyyppiä (Martio 2015, s95-96).

Jos järjestelmä säilyttää hakuhistoriaa, voidaan tämän perusteella arvioida tietokenttien hyödyllisyyttä. Tosin on huomattava, että tietokenttien käyttö haussa ei välttämättä kerro tietokenttien merkityksestä tietomallin kannalta. Muutostapauksissa ylimääräisten tietokenttien tunnistaminen ja poistaminen on tärkeää, jotta tietomalli pysyy selkeänä.

Datan laadun arviointi saatavuuden kautta on hyvin suoraviivaista. Tämä voidaan toteuttaa laskemalla, kuinka suuri osa tietomallin kentistä on täytetty hakukelpoisilla arvoilla.

Jatkuvuutta puolestaan on vaikea arvioida laskemalla. Jatkuvuuden ongelmat ilmenevät tietokenttien epäjohdonmukaisuuksina, joiden lukumäärää ja/tai esiintymistiehyttä voidaan pitää jatkuvuuden mittarina. Epäjohdonmukaisuuksia voidaan olettaa esiintyvän pääsääntöisesti vapaalla tekstillä täytettävissä kentissä. Näin ollen epäjohdonmukaisuuksia voidaan lähteä etsimään näistä tietokentistä. Martio (2015, s95-96) toteaa, että nimikkeiden hyvään hallintotapaan kuuluu, että nimikkeiden kuvauksiin ja spesifikaatioihin on määritelty yrityskohtainen standardoitu esitystapa. Tietomalliin on usein kehitettävä myös datasääntöjä. Epäjohdonmukaisuuksia etsiessä ensin valitaan tarkasteltava tietokenttä, tunnistetaan siihen vaikuttavat säännöt sekä standardiesitystapa, jonka jälkeen tarkastellaan, kuinka suuri osa nimikkeistä ei noudata sääntöä tai esitystapaa. Jos organisaatiolla on käytössä useampia järjestelmiä, jotka hyödyntävät samaa dataa, on myös huomioitava näiden järjestelmien välillä vallitsevat epäjohdonmukaisuudet.

Datan varsinaisien virheiden määrää ei pystytä laskemaan suoraan. On kuitenkin tärkeää, että havaittuihin virheisiin puututaan heti. Datalle määritellyn omistajan rooli nousee tässä yhteydessä esiin, koska käyttäjien on tiedettävä tai löydettävä tieto nopeasti, kenelle ilmoittaa virheestä, jos käyttäjä ei pysty sitä itse korjaamaan. Useamman järjestelmän kokonaisuuksissa on huomioitava, onko sama virhe päässyt kulkeutumaan myös muihin järjestelmiin.

Koska vapaat tekstikentät ovat alttiita epäjohdonmukaisuuksille, datan kestävyydestä antaa viitettä se, kuinka suuri osa tietomallin kentistä on täytettävissä lista-arvoilla vapaan tekstin sijaan. Myös se, että mahdollisimman suuri osa tietokentistä täytetään lista-

arvoilla vapaan tekstin sijaan, kuuluu Martion (2015, s95-96) mukaan hyvään nimikkeiden hallintotapaan.

### **Tietojärjestelmä**

Tietojärjestelmään kohdistuvat muutospaineet tulevat liiketoimintaprosessista. Tietojärjestelmän kyvystä mukautua helposti muuttuviin liiketoiminnan tarpeisiin, eli sen joustavuudesta, antaa viitettä keskimääräisen muutoksen kesto ja kustannukset. Jos esimerkiksi järjestelmän joustavuus on erityisen hyvä, muutosten tekeminen saattaa onnistua nopeasti organisaation tietojärjestelmistä vastaavilta henkilöiltä. Sen sijaan, jos joustavuus on heikkoa, organisaatio saattaa tarvita ulkopuolista apua järjestelmän muokkamiseen, jolloin muutoksen kesto ja kustannukset nousevat merkittävästi.

Tietojärjestelmä toimii datan varastona sekä käsittelyn työkaluna. Näin ollen tietojärjestelmän hakuominaisuudet ovat erittäin merkittävässä roolissa. Hakuominaisuuksien toimintaa on kuitenkin hankala tarkastella muuten kuin datanäkökulman kautta haettavien kenttien täyttöasteena ja lukumääränä. Sen sijaan hakua haittaavana tekijänä voidaan katsoa olevan liian suuri datan määrä, tarkoittaen tässä ylimääräistä, vanhentunutta ja tarpeetonta dataa. Saatavuuden mittarina voidaan pitää tällaisen tarpeettoman datan osuutta kaikesta datasta. Koska tarpeeton data saattaa aiheuttaa lisäkustannuksia sekä kuormittaa järjestelmää, organisaation on hyvä pyrkiä poistamaan tarpeeton data tai vähintään piilottamaan se peruskäyttäjiltä, jolloin se ei häiritse datan hakua.

Tietojärjestelmien kannalta yksi tärkeimmistä datan laatuun vaikuttavista tekijöistä on kullekin datatyypille määriteltävä pääjärjestelmä. Tiettyä datatyyppiä ylläpidetään vain tässä yhdessä pääjärjestelmässä, johon kirjataan dataan kohdistuvat muutokset. Pääjärjestelmästä muutokset replikoidaan muihin järjestelmiin. (Silvola 2018, s40)

Jatkuvuuden erityispiirteenä on vielä duplikaattinimikkeet. Kuten luvussa 3 kävi ilmi, duplikaatit ovat tarpeetonta dataa, joilla on erityisen suuri potentiaali aiheuttaa ongelmia ja lisäkustannuksia. Tästä syystä duplikaattinimikkeiden osuus kaikesta datasta on huomioon otettava mittari.

Tietojärjestelmät mahdollistavat useimmiten datan tehokkaan seurattavuuden. Datan luomisesta ja muokkaamisesta jää järjestelmään jälki, minkä lisäksi järjestelmä tallentaa datan historiatietoa. Datan laadun tekijänä luotettavuus tietojärjestelmänäkökulmasta toteutuu useimmissa järjestelmissä automaattisesti.

Toinen tietojärjestelmien usein mahdollistama tekijä on käyttöoikeuksien hallinta. Käyttäjillä ei tule olla oikeuksia muokata dataa, joka ei heille kuulu. Myös oikeutta datan poistamiseen on rajoitettava mahdollisimman vähäiseen määrään käyttäjistä. Erityisesti PLM-järjestelmät mahdollistavat tehokkaan muutoksenhallinnan, jossa lähtökohtaisesti yksittäisen käyttäjän tekemä muutos ei toteudu ilman, että toinen käyttäjä tarkastaa muutoksen. Nimikkeiden muutoksille määritelty muutoksenhallinta on yksi Martion (2015, s95-96) esittämistä kriteereistä hyvälle nimikkeen hallintotavalle. Näillä toimilla vähennetään merkittävästi virheellisten toimintojen mahdollisuutta. Lisäksi tietojärjestelmän turvallisuuteen kuuluu, että asiattomat eivät pääse dataan käsiksi, mutta tämä menee datan laadun tarkastelun ulkopuolelle.

## Käyttäjä

Loshin (2009) sanoo, että menestyksekkään master datan hallinnan tärkein tekijä ei ole menetelmät tai teknologia, vaan ihmiset. Tarkkuuden ja saatavuuden tasolla datan on oltava oikean henkilöstön käytettävissä. Käyttäjien on osattava tulkita dataa oikein sekä käyttää tietojärjestelmän hakutoimintoja löytääkseen dataa. Datan tarkkuutta voidaan arvioida nimikekohtaisesti tarkastelemalla pystyvätkö käyttäjät nimikkeen tietojen perusteella sanomaan, mikä todellisuuden asia on kyseessä. Tämä tarkastelu tosin on hyvin riippuvainen käyttäjän tiedoista. On myös huomattava, että master datalla on useita käyttökohteita ja käyttäjäryhmiä. Kaikki käyttäjät eivät käytä kaikkia tietokenttiä. Tästä johtuen käyttäjät voivat kokea osan tietokentistä vähemmän tärkeiksi tai jopa tarpeettomiksi. Tietomalli onkin koulutettava kokonaisuudessaan käyttäjille, jotta jokainen tietää, mihin kutakin kenttää käytetään ja mihin se vaikuttaa.

Kuten aiemmin on todettu, dokumentteja käsiteltäessä sisään- ja uloskirjaukset järjestelmästä on toteutettava käyttäjän näkökulmasta mahdollisimman helpoiksi (Martio 2015, s100). Käyttäjänäkökulmasta saatavuutta voidaankin mitata datan hakemiseen käytetyn odotetun ja todellisen ajan erotuksena (Silvola 2018). Tällöin saadaan arvio tietojärjestelmän ja hakutoimintojen käytettävyydestä ja edelleen datan saatavuudesta. Odotusaika riippuu tilanteesta, mutta lähtökohtaisesti odotusaikana voidaan käyttää kokeneen käyttäjän hakuun kuluttamaa aikaa tai jotakin sovittua kiinteää aikaa, jossa haku pitäisi pystyä suorittamaan.

Silvola (2018, s34) mainitsee liiketoimintaprosessien hallinnan yhdeksi tehtäväksi tavoitteiden viestimisen henkilöstölle niin, että kukin organisaatiotaso ymmärtää niiden merkityksen ja niiden vaatimat muutokset. Vastaava tehtävä on myös datan laadun hallin-

nassa. Datan laatua hallinnoivan tahon on viestittävä datan käsittelyyn ja tietojärjestelmiin liittyvät tavoitteet ja muutokset muulle organisaatiolle niin, että kukin henkilöstön jäsen ymmärtää sekä suuremman kuvan datan ja tietojärjestelmien tavoitteisiin liittyen että oman roolinsa tässä suunnitelmassa. Tämän selvittämiseen voidaan hyödyntää ajoittaista haastattelua.

Eppler (2003) on osoittanut tutkimuksessaan, että datan laadussa on subjektiivinen puoli, jossa datan laadukkuus määräytyy sen mukaan, täyttääkö data siihen kohdistuvat odotukset. Tämä muodostaa datan maineen ja uskottavuuden, jotka puolestaan vaikuttavat siihen, missä määrin dataan luotetaan ja tukeudutaan. Datan mainetta ja uskottavuutta voidaan mitata esimerkiksi datan käyttäjille suunnatun kyselyn avulla.

Kaikella datalla kuuluu olla omistaja, joka vastaa datasta. Kun omistaja on määritelty, käyttäjät tietävät kenen puoleen kääntyä dataan liittyvien ongelmien ja kysymysten kanssa. Datan omistajuutta voidaan arvioida datatyyppi kerrallaan yksinkertaisesti tarkastelemalla, onko omistaja määritelty vai ei.

## Prosessi

Prosessin näkökulmasta keskeisin tarkasteltava yksittäinen tekijä on datan tuoma arvo prosessille. Dataan kohdistuvien tarpeiden määrittely edellyttää liiketoimintaprosessien ja -strategian ymmärtämistä (Silvola et al. 2018). Liiketoimintaprosessit voidaan jakaa kahteen luokkaan: strategisesti tärkeisiin ydintoimintoihin ja tukitoimintoihin (Hannus 2004). Ydintoiminnot ovat liiketoimintaprosesseja, jotka näkyvät asiakkaille, kuten myynti, markkinointi, logistiikka ja palvelut. Tukitoimintoja puolestaan ovat muun muassa hankinta, tuotekehitys sekä henkilöstöhallinta. (Silvola 2018, s33) Datan laatu on tärkeässä roolissa kaikissa liiketoimintaprosesseissa, mutta erityisesti ydintoimintojen datan laadusta on huolehdittava, koska niiden kautta data välittyy asiakkaille. Heikkolaatuinen data ydintoiminnoissa heikentää organisaation mainetta asiakkaiden keskuudessa.

Datan laatua prosessin näkökulmasta on erityisen vaikea arvioida. Tarkasteltaessa datan laatua prosessien näkökulmasta, on tärkeää ottaa selvää ja pyrkiä vastaamaan kaikkien liiketoimintaprosessien tarpeisiin. Tällöin usein paras vaihtoehto datan laadun parantamiseen on käyttäjien haastattelu.

Prosessi on myös näkökulmista se, joka määrää dataan kohdistuvan muutospaineen. Prosessorientoituneille organisaatioille, joiden tavoite on reagoida nopeasti muutoksiin, erityinen haaste on ennakoida lähitulevaisuutta. Parhaimmillaan datan hallinnassa onnistunut tulevaisuuden tarpeiden huomioon ottaminen mahdollistaa koko organisaation nopean ja tehokkaan liiketoiminnan reagoinnin. (Silvola 2018, s32)

Jotta organisaatio pysyy kilpailukykyisenä, liiketoimintaprosessien on muututtava jatkuvasti kilpailijoiden ja ympäristön mukaan (Silvola 2018, s32). Näin ollen myös datan hallinnassa on seurattava jatkuvasti muuttavia liiketoiminnan tarpeita. Jos näin ei tehdä, datan laadun voidaan katsoa heikkenevän jatkuvasti, datan laadun ja sen hallinnan erkaantuessa liiketoiminnan todellisista tarpeista. Tätä kutsutaan datan rappeutumiseksi (engl. *data decay*). Datan rappeutumisen vuoksi liiketoimintaprosessien ja -strategian dataan kohdistamat tarpeet on kartoitettava säännöllisesti haastatteleamalla asianmukaisia henkilöitä.

Viimeinen prosessinäkökulman datan laadun osatekijöistä on turvallisuus, joka ilmenee käyttövarmuutena. Käyttövarmuutta voidaan mitata merkittävien käyttökatkosten lukumääränä vuotta kohden.

Pääsääntönä voidaan sanoa, että datan laadun kartoittaminen data- ja tietojärjestelmänäkökulmien suhteen vaatii edellä esitettyjen mittarien ylläpitoa sekä datan ja tietojärjestelmien asiantuntevaa tarkastelua. Käyttäjä- ja prosessinäkökulmat sen sijaan vaativat säännöllistä keskustelua eri käyttäjien ja vastuutahojen kanssa. Esitetyille mittareille on hyvin vaikeaa sanoa kaiken kattavia raja-arvoja. Esimerkiksi duplikaattien osuutena kaikki yli 0 % on huonoa ja datan laatua heikentävää. Raja-arvoja tärkeämpänä voidaan pitää trendejä, jotka paljastuvat säännöllisistä mittauksista.

Datan laadulle ei voida osoittaa tarkkoja kehittämistoimenpiteitä. Kaikkien mittausten ja haastattelujen pohjalta tunnistettujen ongelmakohtien kanssa on pyrittävä löytämään ongelman pohjalla piilevät syyt (Loshin 2011). Tutkimukset osoittavat, että lähes puolet organisaatioista tyytyvät vain korjaamaan datan laatuongelman sen sijaan, että korjaisivat ongelman aiheuttajan (Russom 2006). Jos mittari esimerkiksi osoittaa tietokenttien vajaan täyttöasteen, kenttien täyttäminen jälkikäteen korjaa ongelman hetkellisesti. Se ei kuitenkaan korjaa ongelman syytä, joka voi olla esimerkiksi käyttäjien huolimattomuus tietokenttien täytön suhteen. Ongelmille voi olla lukuisia syitä, mutta yleisimpinä ratkaisuina voidaan mainita käyttäjien koulutus sekä tietojärjestelmien konfigurointi niin, että ne ohjaavat käytössä pois datan laadun virheistä ja ongelmista. Tunnistamalla ongelmien taustalla olevat syyt voidaan siirtyä reaktiivisesta toiminnasta proaktiiviseen lähestymistapaan (Loshin 2011). Kun edellisessä luvussa esitettyyn taulukkoon 3.5 lisätään tässä luvussa esitetyt datan laadun mittarit ja testaustavat, saadaan taulukko 4.2.



**Taulukko 4.2. Datan laadun viitekehys mittarien ja testaustapojen kanssa.**

|  | <b>Data</b>   | <b>Tietojärjestelmä</b>  | <b>Käyttäjä</b>  | <b>Prosessi</b>   |
|--|---|--|--|---|
| <b>Tarkkuus</b><br>Datalla pystytään tietystä tietojärjestelmä kontekstissa kuvaamaan ymmärrettävästi relevanttia todellisuutta riittäväällä tarkkuudella ja laajuudella.  | Tietomallin toimivuus<br><br>Kartoitus: haastattelu   | Joustavuus<br><br>Kartoitus: keskimääräisen muutoksen kesto ja kustannukset  | Tulkittavuus<br><br>Kartoitus: nimikkeen peilaus todellisuuteen nimiketietojen perusteella | Relevanttius<br><br>Kartoitus: haastattelu                                |
| <b>Saattavuus</b><br>Data on tietomallissaan kattavaa ja näin ollen helposti haettavissa ja käytettävissä tietojärjestelmästä.   | Kattavuus<br><br>Kartoitus: tietokenttien täyttöasteet                                      | Haun toimivuus ja datan määrä<br><br>Kartoitus: tarpeettoman datan osuus   | Helppokäyttöisyys<br><br>Kartoitus: hakuun käytetty aika                                   | Oikea data oikeassa paikassa oikeaan aikaan<br><br>Kartoitus: haastattelu |
| <b>Jatkuvuus</b><br>Data ei sisällä epäjohdonmukaisuuksia, eikä sillä ole epäjatkuvuuskohtia eri prosessien ja järjestelmien välillä, eikä data näin ollen aiheuta prosessille ylimääräisiä kustannuksia tai ongelmia. | Johdonmukaisuus<br><br>Kartoitus: keskimääräinen epäjohdonmukaisuuksien osuus tietokentistä | Integraation toimivuus ja duplikaattomuus<br><br>Kartoitus: pääjärjestelmä määritelty: kyllä/ei. Duplikaattinimikkeiden osuus                    | Selkeys<br><br>Kartoitus: haastattelu  | (Kustannus)tehokkuus<br><br>Kartoitus: haastattelu                        |
| <b>Luotettavuus</b><br>Käyttäjä voi luottaa tietojärjestelmästä saadun datan olevan käyttökelpoista ja paikkansa pitävää.  | Virheettömyys, objektiivisuus<br><br>Kartoitus: -   | Seurattavuus<br><br>Kartoitus: historiatieto tallennetaan: kyllä/ei  | Maine, uskottavuus<br><br>Kartoitus: kysely  | Ajantasaisuus<br><br>Kartoitus: haastattelu                               |
| <b>Turvallisuus</b><br>Kerran laadukkaaksi todettu data ei herkästi muutu heikkolaatuiseksi.   | Kestävyys<br><br>Kartoitus: lista-arvoja käyttävien kenttien osuus tietomallissa            | Käyttöoikeudet ja muutoksenhallinta<br><br>Kartoitus: käyttöoikeuksien tarkastaminen käyttäjäryhmittäin. Muutoksenhallinta määritelty: kyllä/ei. | Omistajuus<br><br>Kartoitus: omistaja määritelty: kyllä/ei                                 | Käyttövarmuus<br><br>Kartoitus: käyttökatkoksien määrä vuodessa           |

## 5. TUTKIMUSMENETELMÄ

Tässä luvussa esitellään työssä käytetty tutkimusmenetelmä, empiirisen aineiston kerääminen sekä tutkimuksen eteneminen. Lisäksi luvussa esitellään työn kohdeyrityksen lähtötilanne sekä tutkimuksen sijoittuminen ajallisesti kohdeyrityksen tietojärjestelmä-uudistuksessa.

### 5.1 Tapaustutkimus

Työn tutkimus toteutetaan kvalitatiivisena tapaustutkimuksena, jossa tarkastellaan yhtä kohdeyritystä. Tutkimus keskittyy aihealueen tarkasteluun kohdeyrityksen kontekstissa. Tapaustutkimus mahdollistaa valitun kohteen syvällisen tarkastelun ja auttaa lisäämään ymmärrystä aiheesta (Meredith 1998). Tapaustutkimuksella pystytään hyvin tuottamaan vastauksia miten ja miksi kysymyksiin (Yin 2008). Kohdeyrityksen näkökulmasta tapaustutkimus tuottaa erityisen olennaisia tuloksia johtamisen kannalta (Meredith 1998).

Tapaustutkimukset ovat usein työläitä toteuttaa. Tapaustutkimuksen luonne vaatii monesti useamman menetelmän yhtäaikaista käyttöä sekä menetelmien välistä vuoropuhelua. (Meredith 1998) Lisäksi yhden tapauksen tarkastelussa on monitapaustutkimuksia suurempi riski tulosten vinoutumiselle. Tulokset eivät välttämättä myöskään ole yleistettävissä yhden tapauksen perusteella. (Voss et al. 2002) Työssä ei kuitenkaan pyritä yleistettävään tietoon vaan tuottamaan nimenomaiselle yritykselle kohdistettua tietoa.

Yin (2008) listaa viisi yhden tapauksen analyysityyppiä. Ensimmäinen tyyppi on kriittinen tapaus, jossa painottuu tapauksen valinta tutkimuskysymykseen sopivaksi. Toinen tyyppi on ainutlaatuinen tapaus, jossa tarkasteltava ilmiö esiintyy äärimuodossaan. Kolmas tyyppi on paljastava tapaus, jossa uutta ilmiötä kuvataan varhaisessa vaiheessa. Neljäs tyyppi, tyypillinen tapaus, edustaa normia. Viimeinen tyyppi on pitkittäistapaus. Pitkittäistapauksessa tutkitaan muutoksia yli ajan. Tämän työn tutkimuksen voidaan katsoa edustavan kriittistä tapausta, koska kyseiselle yritykselle ei voida tuottaa yhtä hyödyllistä tietoa tarkastelemalla jotakin muuta yritystä. Toisaalta työn tutkimus edustaa osaltaan myös pitkittäistapausta, sillä tarkastelu jakautuu menneisyyteen, nykyhetkeen ja tulevaisuuteen.

Lincoln & Cuba (1985) määrittelevät kvalitatiiviselle tutkimukselle viisi laatukriteeriä, jotka ovat uskottavuus, siirrettävyys, luotettavuus, varmennettavuus ja refleksiivisyys. Uskottavuus kuvaa sitä, kuinka paljon tutkimuksen tuloksiin voi luottaa ja ovatko tulokset oi-

keita. Korkea siirrettävyys tarkoittaa, että tulokset ovat hyvin siirrettävissä toiseen kontekstiin tai toisille testihenkilöille. Luotettavuus kuvaa tulosten stabiiliutta ja sitä, missä määrin tuloksista muodostetut johtopäätökset ja tulkinnat ovat tutkimuksen tulosten tuomia. Varmennettavuus tarkoittaa muiden tutkijoiden mahdollisuutta toistaa tutkimus ja saada samat tulokset. Viimeisenä laatukriteerinä refleksiivisyys kuvaa sitä, miten hyvin tutkimuksen tekijä on onnistunut huomioimaan tutkimuksessa ja tuloksissa oman subjektiivisuutensa ja suhteensa muihin tutkimuksessa olleisiin henkilöihin. (Lincoln & Cuba 1985) Näihin laatukriteereihin palataan myöhemmin pohdinnassa tutkimuksen rajoitteista.

## 5.2 Kohdeyritys

Työn kohdeyritys on valmistavan teollisuuden pk-yritys. Yritys on ollut olemassa työn kirjoitushetkellä jo lähes 10 vuotta, joista suurin osa on käytetty tuotekehitykseen. Yrityksessä työskentelee n. 30 henkilöä. Yrityksellä on yksi tuotantolaitos Suomessa sekä kaksi tuotantolaitoksesta erillistä toimipistettä, joista toinen sijaitsee Suomessa ja toinen Yhdysvalloissa. Yritys on aloittanut myynnin vasta viime vuosina. Markkinoiden kiinnostuneisuus yritystä kohtaan, teknologiamurros sekä kilpailukykyinen tuote ennakoivat yritykselle merkittävää kasvupotentiaalia. Yrityksen johto tiedostaa tämän kasvupotentiaalin sekä myös yrityksen tulevaisuudessa kohtaamat haasteet, mikäli se jatkaisi nykyisellä tiedonhallinnan toteutuksella ilman varsinaisia tietojärjestelmiä.

Yrityksen dataa hallitaan lähtötilanteessa pääsääntöisesti Excelillä. Data on tallennettuna useisiin Excel-tiedostoihin, jotka puolestaan ovat tallennettuna yrityksen verkkolevyille. Verkkolevy mahdollistaa datan saatavuuden kaikille käyttäjille. Yrityksen sisäiset rajoitukset datan käytölle on toteutettu yksinkertaisesti jakamalla yksittäiset verkkolevyille tallennetut kansiot vain niitä tarvitseville henkilöille. Excel-datan lisäksi yrityksellä on runsaasti CAD-suunnitteludataa sekä dokumentteja. Myös suunnitteludata ja dokumentit ovat tallennettuna verkkolevyille.

Kohdeyrityksen keskeinen tavoite on vahva globaali kasvu. Tämän vuoksi on erittäin tärkeää luoda yritykseen toimiva ja ennen kaikkea skaalautuva tietojärjestelmäarkkitehtuuri, joka tukee toimintaa sekä lyhyellä tähtäimellä, yrityksen ollessa vielä suhteellisen pieni, että pitkällä tähtäimellä, yrityksen kasvaessa. Tällä hetkellä esimerkiksi tuotehallinnan toteuttaminen Excel-taulukoilla on huonosti skaalautuva toimintatapa johtuen muun muassa yhteiskäytön puutteesta ja käytön liiallisesta vapaudesta.

Kohdeyrityksessä toteutettiin tietojärjestelmäprojekti, jossa yrityksessä otettiin käyttöön ERP- ja PLM-järjestelmät. ERP-järjestelmäksi valittiin Odoo ja PLM-järjestelmäksi Aras

Innovator. Järjestelmät integroidaan toisiinsa siten, että PLM-järjestelmä toimii tuotetiedon pääjärjestelmänä. Diplomityön lähtötilanteessa molemmat järjestelmät on kilpailutettu, testattu ja valittu. Molemmat järjestelmät otetaan käyttöön moduuleittain koko organisaatiossa eli implementointitapa on kuvan 2.2 vaiheittainen implementointi organisaatiotasolla (vasen yläkulma). Työn tutkimuksen aikaan PLM-järjestelmä on ensimmäisessä stabilisointivaiheessaan ja ERP-järjestelmä implementointivaiheessa. Diplomityön loppuvaiheessa PLM-järjestelmä on edennyt käyttövaiheeseen, kun taas ERP-järjestelmä puolestaan on stabilisointivaiheessa. Molemmista on myös käynnistetty seuraava implementointivaihe.

### 5.3 Empiirisen aineiston kerääminen

Vaikka datan laadun kvantitatiiviselle analyysille on yleisesti tunnistettu tarve (Wand & Wang 1996), työn empiirinen aineisto kerättiin kohdeyrityksen henkilöstöä haastattelella. Datan laadun kvantitatiivinen tarkastelu ei ollut kohdeyrityksen tapauksessa järkevää, koska tutkimus tehtiin tietojärjestelmä uudistuksen aikana. Tietojärjestelmiä päivitettiin jatkuvasti, niiden käyttöä opeteltiin ja migraatiossa tietojärjestelmiin ajettua dataa käsiteltiin. Näistä syistä kvantitatiivinen data ei olisi ollut riittävän luotettavaa ja se olisi tuottanut harhaan johtavia johtopäätöksiä. Tämän vuoksi työn varsinaisessa tutkimuksessa pyrittiin kartoittamaan datan laatua mahdollisimman laajasti haastattelujen avulla.

Haastatteluja tehtiin kahdeksan kappaletta. Haastattelut toteutettiin noin kolme kuukautta PLM-järjestelmän käyttöönoton jälkeen ja ennen ERP-järjestelmän käyttöönottoa. Koska tuleva ERP-järjestelmä integroitui käyttämään PLM-järjestelmän dataa, oli erityisen tärkeää havaita mahdolliset datan laatuongelmat ennen järjestelmien integroimista.

Haastattelut olivat puolistrukturoituja. Haastatteluissa keskustelu pyrittiin pitämään mahdollisimman vapaana aihealueen puitteissa, jotta saataisiin mahdollisimman laaja ymmärrys kohdeyrityksen tilanteesta. Haastatteluissa pyrittiin muodostamaan useita, jopa hieman subjektiivisia, näkemyksiä kohdeyrityksen datan laadusta ennen tietojärjestelmä uudistusta, uudistuksen aikana sekä tulevaisuudessa. Haastatteluissa käytetty kysymysrunko on esitetty liitteenä A. Haastattelut äänitettiin ja ne kestivät keskimäärin noin 20-30 minuuttia.

Haastatteluihin pyrittiin valitsemaan työssään runsaasti master dataa hyödyntäviä henkilöitä. Haastateltavien valinnassa otettiin myös huomioon erilaiset lähestymistavat master dataan, millä pyrittiin saamaan tuloksiin useita erilaisia näkökulmia. Haastatellut henkilöt ovatkin jaettavissa master dataa tuottaviin, käyttäviin ja hallinnoiviin tahoihin. Vastaavaa jakoa ovat käyttäneet myös esimerkiksi Strong et al. (1997). Erona Strong et al.

(1997) käyttämään luokitteluun, datan hallinnoijiin on tässä luettu mukaan myös datan luomista johtajat ja ohjaavat tahot, kuten tuotepäällikkö. Strong et al. (1997) sisällyttää hallinnoivaan ryhmään vain tietoteknisiä resursseja, kuten datan säilytystä ja prosessointia, tarjoavat ja hallinnoivat tahot. Tässä työssä näitä tahoja edustaa tietojärjestelmäpäällikkö. Toisaalta on huomattava, että useimmat haastatelluista henkilöistä toteuttavat useampaa näistä rooleista. Taulukossa 5.1 on esitetty haastateltujen työnkuvat, luokittelu käyttäjiin, tuottajiin ja hallinnoijiin, haastattelujen kestot, sekä nimimerkit, joilla heihin tuloksissa viitataan.

**Taulukko 5.1.** Työhön haastateltujen työnkuvat sekä nimimerkit.

| Haastattelun numero | Haastateltavan luokittelu | Haastateltavan työnkuva   | Haastattelun kesto (minuutit : sekunnit) | Nimimerkki    |
|---------------------|---------------------------|---------------------------|--|---------------|
| 1                   | Hallinnoija               | Tuotepäällikkö            | 29:54                                    | TuoteP.       |
| 2                   | Hallinnoija               | Tietojärjestelmäpäällikkö | 25:50                                    | Tietojärj.P.  |
| 3                   | Hallinnoija               | Tekninen johtaja          | 51:59                                    | Tek.J.1       |
| 4                   | Hallinnoija               | Tekninen johtaja          | 23:01                                    | Tek.J.2       |
| 5                   | Käyttäjä                  | Talouspäällikkö           | 25:47                                    | TalousP.      |
| 6                   | Käyttäjä                  | Hankintavastaava          | 35:54                                    | HankintaV.    |
| 7                   | Käyttäjä                  | Varastovastaava           | 23:39                                    | VarastoV.     |
| 8                   | Tuottaja                  | Suunnittelija             | 22:58                                    | Suunnittelija |

Haastateltujen määrä edustaa n. 25 % koko kohdeyrityksen henkilöstöstä, joten määrän voidaan katsoa olevan edustava. Taulukossa 5.1 esitetään, että master datan tuottajiksi luokiteltuja henkilöitä on vain yksi. Master datan hallinnoijiksi luokitellut tekniset johtajat ovat kuitenkin myös master datan käyttäjiä ja tuottajia, joten nämä näkökulmat tulevat heidän myötensä riittävän kattavaksi. Kohdeyrityksen pienestä koosta johtuen osa haastatteluihin valituista henkilöistä tekee myös työnkuvassaan usean toiminnon töitä. Esimerkiksi haastateltu varastovastaava työskentelee myös tuotannossa. Myös tästä syystä työhön saatiin laaja näkökulma, eikä lisähaastatteluille ollut tarvetta.

## 5.4 Empiirisen aineiston analysointi

Kaikki haastattelut nauhoitettiin ja niistä litteroitiin tekstinkäsittelyohjelmalla asiasisällöltään ja relevanttiudeltaan oleelliset osat. Tässä vaiheessa suodatettiin pois kontekstiin liittymättömät asiat, asiasisällöltään heikot asiat, kuten tautologiat, sekä epärelevantit puhekieliset ilmaukset ja täytteet.

Osat jaoteltiin ensin haastattelurungon mukaisesti käsittelemään ajankohtaa ennen tietojärjestelmä uudistusta, sen aikana tai sen jälkeen ja kaikkien haastattelujen saman ajankohdan litteroinnit kerättiin yhteen. Tämän jälkeen litteroinneista muodostettiin selkeitä, pääasiassa yhtä aihealuetta tai datan laadun osatekijää kuvaavia kommentteja. Kommentit ryhmiteltiin aiheen perusteella siten, että samaa aihetta kuvaavat kommentit olivat peräkkäin.

Kommenttien pohjalta muodostettiin listaus kohdeyrityksen keskeisimmistä datan laatuun liittyvistä ongelmista, jota käsitellään luvussa 7. Kommenteista kerättiin myös tilasto kunkin datan laadun osatekijän esiintymismäärästä, minkä jälkeen selvät päällekkäisyydet poistettiin.

Lopuksi kommentit pyrittiin järjestelemään aihealueittain mahdollisimman selkeästi eteneväksi kokonaisuudeksi. Kommentteja myös siistittiin vähemmän puhekieliseen muotoon, jotta itse asiasisällön seuraaminen olisi helpompaa.

## 6. TULOKSET

Tässä luvussa käydään läpi haastatteluiden tulokset. Haastatteluissa nousi esiin paljon erilaisia näkemyksiä kohdeyrityksen datan laadusta. Suurinta hajontaa esiintyi näkemyksissä datan laadusta ennen tietojärjestelmäuudistusta. Haastatteluissa korostui odotetusti hyvin vahvasti kunkin haastateltavan toimenkuva ja rooli datan suhteen (käyttäjä, tuottaja tai hallinnoija).

Tämän luvun ensimmäisessä osassa tarkastellaan, miten haastateltavat ymmärsivät datan laadun yleisesti. Toisessa osassa käydään läpi, millaisia ongelmia datan laadussa havaittiin ennen tietojärjestelmäuudistusta. Kolmas osio käsittelee nykyhetkeä ja sitä, millaisia datan laatuun liittyviä havaintoja on tehty tietojärjestelmäuudistuksen aikana. Viimeinen luku kartoittaa haastateltavien näkemyksiä datan laadusta tulevaisuudessa.

Luvun siteerauksissa puhutaan sekä CAD-tiedostoista että SolidWorksista. SolidWorks on kohdeyrityksen CAD-ohjelmisto ja CAD-tiedostoilla viitataan SolidWorksillä tuotettuihin tiedostoihin.

### 6.1 Näkemys datan laadusta ja siihen suhtautuminen

Datan laatu koettiin laajasti erittäin tärkeäksi ja datan laadun merkitys omaan työhön koettiin keskeiseksi. Hyvälaatuisen datan katsottiin parantavan myös työn laatua.

*”Mitä parempaa dataa niin sen parempaa työn jälkeä, sen helpompi on tehdä työtä.” (Suunnittelija)*

Joissakin haastatteluissa korostui datan rooli yrityksen liiketoiminnassa eli prosessinäkökulma. Data nähtiin keskeisenä osana yrityksen toimintaa ja toimintakykyä.

*”Hyvälaatuinen data on mun mielestä semmosta joka tukee sitä bisneksen liiketoimintaa. Sieltä saa helposti sitä tietoa pihalle mitä haluaa.” (Tietojärj.P.)*

*”Dataa tullaan kysyyn meiltä tosi paljon enemmän kuin aikaisemmin ja sen pitää olla laadukasta. Ihan varmasti määrällisesti sen tarve tulee räjähtään ja sitten sen pitää olla laadultaan hyvää.” (TalousP.)*

Laadukkaan datan yhtenä keskeisimpänä hyötynä nähtiin jokapäiväisen työn helpottaminen ja ylimääräisen työn välttäminen mikä liittyy myös prosessinäkökulmaan. Osatekijöistä nousi tässä yhteydessä esille erityisesti tarkkuus ja saatavuus eli datan on kuvattava riittävällä tarkkuudella todellisuutta ja datan on löydyttävä oikeasta paikasta. Datan saatavuuden suhteen on huomioitava, että datan on oltava saatavilla prosessinäkökulmassa kaikille niille yrityksen toiminnoille, jotka kyseistä dataa käyttävät.

*”...laadukkaalla datalla säästetään aika paljon ylimääräistä työtä ja jos se on laadukasta, niin asiat selviää huomattavasti nopeammin kuin huonolaatuisella datalla.” (TuoteP.)*

*”Se [hyvälaatuinen data] löytyy myös niistä paikoista, joissa sitä halutaan käyttää.” (Tietojärj.P.)*

Lisäksi talouspäällikön haastattelussa nousi esiin, että datan laatuun saatetaan kiinnittää huomiota vasta sitten, kun laatu on todettu huonoksi. Tämä viittaa reaktiiviseen lähestymistapaan datan suhteen, jossa ongelmia korjataan vasta, kun niitä esiintyy, ennaltaehkäisyyn sijaan.

*”...master data näkyy itselläkin tosi monessa paikassa silloin kun se on kuraa. Silloin, kun master data ei toimi omasta näkökulmasta, niin se näkyy sitten sellaisena, että joko ei saada sieltä järjestelmistä sitä dataa mitä halutaan tai sitä joudutaan manuaalisesti... uudelleentyöstämään. Tai sitten se, että master data on esimerkiksi puutteellista saattaa aiheuttaa tässä omassa tontissa [taloushallinnossa] jotakin muuta manuaalista työtä jossakin vaiheessa.” (TalousP.)*

Datan laadun osatekijät osattiin nimetä pääsääntöisesti melko hyvin. Kullakin haastateltavalla kuitenkin korostui hänelle itselleen tärkeät tekijät muita enemmän. Tekninen johto piti erityisen tärkeänä sitä, että data kuvaa riittävän hyvin todellisuutta (osatekijöistä tarkkuus datanäkökulmasta). Tällä nähtiin olevan vahva vaikutus ylimääräisen työn välttämiseen tulevaisuudessa.



*”Vähän ylempänä käsitteenä... ymmärtää mikä se komponentti on ja mitkä sen meille kriittiset arvot on... Eli millä kriteereillä me ollaan se valittu ja jos se uudelleen valitaan niin mitkä olisi ne kriteerit myös jatkossa.” (Tek.J.2)*

Esimerkkinä heikkolaatuisen datan aiheuttamasta ylimääräisestä työstä varastovastaa nosti esille jatkuvuuden tietojärjestelmänäkökulman ongelman, jossa tietoa joutuu etsimään useammasta paikasta.

*”[Sen datan], mikä meillä on siellä ERP-järjestelmässä, mä pystyn jäljittään sen osan täydellisesti sitä kautta eikä mun tarvi mennä toiseen dokumenttiin taas metsästämään että mä pystyn löytään jotain tietoa tai valmistajan koodeja.” (VarastoV.)*

Datan virheettömyys mainittiin haastatteluissa useasti, mutta lähinnä vain ohimennen. Se nähtiin kuitenkin yhtenä tärkeänä datan laadun osatekijänä. Luotettavuuden osatekijän datanäkökulma (virheettömyys) johtaa käyttäjänäkökulmaan (datan maine).

*”Datan laatu merkitsee tietysti sitä, että mä voin luottaa siihen, siis etunenässä sitä, että datan laatu on sellasta että se on oikeaa.” (VarastoV.)*

Datan hallinnointiin keskittyvät tahot painottivat osatekijöistä datan jatkuvuutta data- ja tietojärjestelmänäkökulmista. Haastatteluissa korostettiin kokonaisuuden hallintaa sen sijaan, että hallittaisiin yksittäisiä osia.

*”Hyvälaatuinen data on myös semmosta, että se mukailee aina sitä samaa formaattia ja siinä voi luottaa, että sieltä aina löytyy kaikki ne tiedot mitkä on määritetty pakollisiksi tiedoiksi.” (Tietojärj.P.)*

*”[Hyvälaatuinen data] on riittävän laajasti ainakin yhtenäistä, että ei ole niin että... tietyt nimikkeet on äärettömän hyviä ja tietyt nimikkeet on sitten hyvinkin heikkolaatuista. Se taso olisi että voidaan ainakin luottaa siihen mitä on, ja että ne kriittiset arvot, mitkä on oikeesti kriittisiä, on sitten kaikista olemassa.” (Tek.J.2)*

*”... hallitaan koko ajan sitä koko dataa, ettei sitä hallita erikseen osittain.” (Tek.J.2)*

Myös datan loppukäyttäjät tiedostivat datan jatkuvuuden ja dataprosessin kokonaisuuden hallinnan tärkeyden. Esille nousi myös jatkuvuuden käyttäjänäkökulma (selkeys).

*”Se, että kuinka hyvin se prosessi toimii datan näkökulmasta oikein, että se tuottaa oikeanlaista dataa, on yks keskeisiä asioita. ... Sekä järjestelmämielessä että sitten miten ihmiset toimii siellä, minkälainen se prosessi on ja miten sitä järjestelmää käytetään.” (TalousP.)*

*”[Esimerkiksi varastosaldoissa tai rakenteissa] tieto on oikeeta ja niissä ei ole ris-tiriitaisuuksia. Jos samaa dataa, tai saman osan dataa löytyy monesta paikasta, niin se on synkassa keskenänsä, että sieltä ei löydy mitään poikkeavuuksia.” (VarastoV.)*

Ajantasaisuus, datan oikea määrä ja turvallisuus jäivät yllättävän pienelle huomiolle useimmissa haastatteluissa. Nämä tekijät kuitenkin esiintyivät useasti tarkasteltaessa yrityksen aiempia käytäntöjä sekä nykytilaa ja tulevaisuutta.

*”Riittävä määrä tietoa, mutta ei ole liikaa tietoa, jossa ei toisaalta ole tarpeettomasti toistettu.” (VarastoV.)*

Haastateltavilla oli pääsääntöisesti hyvä käsitys siitä, mitä datan laatu on. Taulukossa 6.1 on esitetty datan laadun osatekijöiden kunkin näkökulman esiintymismäärät työn kaikissa haastatteluissa. Taulukossa on lisäksi esitetty jokaisen osatekijän (vaakarivit) ja näkökulman (pystysarakkeet) yhteenlasketut esiintymismäärät.

**Taulukko 6.1.** *Datan laadun osatekijöiden esiintymismäärät haastatteluissa.*

|              | Data | Tietojärjestelmä | Käyttäjä | Prosessi | Yhteensä |
|--------------|------|------------------|----------|----------|----------|
| Tarkkuus     | 11   | 7                | 4        | 4        | 26       |
| Saatavuus    | 4    | 5                | 10       | 5        | 24       |
| Jatkuvuus    | 4    | 10               | 5        | 6        | 25       |
| Luotettavuus | 4    | 8                | 6        | 6        | 24       |
| Turvallisuus | 9    | 11               | 4        | 5        | 29       |
| Yhteensä     | 32   | 41               | 29       | 26       |          |

Kuten taulukosta havaitaan, kunkin osatekijän jokainen näkökulma nousi esiin haastatteluissa. Lukumäärät osoittavat, että useat tekijöistä (esiintymismäärä 4-6 kertaa) mainittiin vain ohimennen tai niitä painottivat vain osa haastateltavista. Selvästi eniten huomiota saivat tarkkuuden datanäkökulma (esiintymismäärä 11), saatavuuden käyttäjänäkökulma (esiintymismäärä 10), sekä jatkuvuuden ja turvallisuuden tietojärjestelmänäkökulmat (esiintymismäärät 10 ja 11). Tämä viestii siitä, että erityisesti näihin tekijöihin on syytä kiinnittää huomiota.

## 6.2 Datan käyttö ja datan laatu ennen tietojärjestelmäuudistusta

Kohdeyrityksen tuotehallinta oli ennen tietojärjestelmäuudistusta toteutettu Excel-tilukoon ja CAD-tiedostoihin. Tämän nähtiin laajasti tuottavan ongelmia, koska Excel-tilukko ja CAD-tiedostot olivat omia kokonaisuuksiaan, jotka eivät automaattisesti keskustelleet lainkaan keskenään. Käyttäjät joutuivat käsin syöttämään jotkin tiedot molempiin. Näin ollen jatkuvuuden tietojärjestelmänäkökulma oli heikolla pohjalla.

*”Osa niistä datoista on ylläpidettäviä molemmissa [Excelissä ja CADissa]. Se oli periaatteessa vähän ongelma. ... nimikenumero on haettu Excelistä ja sen jälkeen sen joutuu manuaalisesti joko copy-pastella tai vaan kirjoittaa sitten [SolidWorksiin] tiedostonimeen. Siinä on se ensimmäinen virheen mahdollisuus...” (TuoteP.)*

Tuotehallinnan käytäntö oli kuitenkin melko suoraviivainen, erityisesti vain pääsääntöisesti Excel-tiedostoa käyttäville. Näin ollen jatkuvuuden käyttäjänäkökulma (selkeys) toteutui tältä osin. On kuitenkin huomattava, että kokonaisuus saattoi näille käyttäjille olla vähemmän selkeä.

*"[Tuotehallinta] sinänsä oli hyvin suoraviivaista eli kaikille nimikkeille, oli ne sitten CAD-nimikkeitä, ostonimikkeitä tai myyntinimikkeitä, käytännössä mitä tahansa, oli Excelissä oma rivi ja sille oli tietyt attribuutit. Se oli hyvin selkee. Tuoterakenne muodostettiin samaan Exceeliin." (TuoteP.)*

Tuotehallinnan Excel oli hyvin monipuolinen, mutta näin ollen myös hyvin monimutkainen, mikä heikensi saatavuutta sekä tietojärjestelmä- että käyttäjänäkökulmasta. Lisäksi Excel-taulukkoa ei nähty hyvin skaalautuvana ratkaisuna yrityksen koon ja nimikkeiden määrän kasvaessa.

*"Aika selkeä ja kattava kylläkin, mutta tottakai siinä rupes tuleen vastaan jo sekin että kun nimikkeiden määrä kasvaa niin tommonen Excelihän käy sitten jo hitaaks ja hankalaks, rivejä ja sarakkeita julmettu määrä." (Suunnittelija)*

Excelissä toteutusta tuotehallinnasta nousi esiin myös haasteet tarkkuuden data- ja käyttäjänäkökulmista. Selvästi tarkkuuden datanäkökulma eli tietomalli ei ollut optimaalinen, mikä aiheutti ongelmia tarkkuuden käyttäjänäkökulmassa eli datan tulkinvaraisuutta.

*"Yksilölliset koodit on ollut käytössä alusta lähtien ja Exceeliä on hallittu tosi taidokkaasti, sillä on pystytty tekemään monenlaisia asioita... Siellä Excel tasolla data on hyvälaatuista, mutta se on jossain määrin vaikeasti käytettävää ja vaikeasti tulkittavaa." (Tietojärj.P.)*

*"Ja sitten on sellaisia kokonaisuuksia, että yhden koodin alla on kokonaisuus ja se ei sitten käy kuitenkaan siitä rivistä ilmi. Siinä on koodi ja siinä kerrotaan mistä on kyse suurin piirtein, mutta se varsinainen sisältö ei selviä mistään." (VarastoV.)*

Samaa Exceliä käytettiin myös dokumenttien hallintaan, mikä aiheutti sekaannuksia. Lähtökohtaisesti komponenttinimikkeiden hallintaan suunniteltu Excel ei myöskään pystynyt taipumaan aina riittävästi dokumenttien vaatimuksiin. Tämä on ongelma tarkkuuden tietojärjestelmänäkökulmasta.

*”Dokumenteista on pidetty kyllä kirjaa, mutta huonona puolena Excel ei ole tuonut kauheasti ulottuvuuksia. Siellä on mennyt vähän sekaisin se käsite, onko kyseessä dokumentti vai onko kyseessä [komponentti] ja liittyykö ne toisiinsa. Ei ole pystytty edes kertomaan niin selkeästi mitkä dokumentit liittyy mihinkäkin [komponentteihin].” (Tietojärj.P.)*

Varsinaisien järjestelmien puute hankaloitti datan hallintaa. Osatekijöistä turvallisuudessa oli näin ollen myös parantamisen varaa, erityisesti data- ja tietojärjestelmänäkökulmista. Esimerkiksi uuden datan validointi oli todella vaikeaa.

*”Vanhassa järjestelmässä [nimikkeen lisääminen] on ollut käytännössä hyvin vapaata; lisäät nimikkeen ja kun oot lisännyt sen, se on olemassa. Ei sitä hyväksytäkseen tai tarkasteta.” (Tek.J.2)*

*”Oli tosi vahvasti se, kun me Exceliä käytetään, jos joku tekee jonkun virheen johonkin, niin se tapahtuu hyvin helposti eikä sitä huomaa. ... meillä on hyvin paljon piirustusvirheitä meidän piirustuksissa, ihan sellasia yksinkertasia jota tapahtuu jokaiselle hyvinkin usein, mutta sitten kun ei ole minkäänäköistä tarkastusporttia siinä välissä, suunnittelija pystyy sen piirtämään ja samalla minuutilla me pystytään hankkimaan se. On käynyt paljon näin että piirustuksissa on ollut virheitä ja se on kumuloinut sitten taas paljon lisätyötä.” (HankintaV.)” (HankintaV.)*

*”Sinänsä siinä [Excelissä] asiat ihan ok, mutta siinä äkkiä virheiden mahdollisuus kasvaa... siinä voi tulla inhimillisiä virheitä ihan tiedon syötössä tai tiedon luvussa; katsoo väärästä sarakkeesta tai vierestä.” (Suunnittelija)*

Tietojärjestelmä uudistusta edeltäneestä tilanteesta nousi erittäin selvästi esille, että järjestelmien puute ja osittain hajanainen datan hallinta jätti datan laadun käytännössä täysin datan tuottajien vastuulle. Tämän todettiin haastatteluissa aiheuttaneen useita puutteita sekä ongelmia erityisesti jatkuvuuden kannalta.

*”On jollain tyylillä sovittu jossain ryhmässä, että ruvetaan laittaa kuvanumeroita... Mutta toisaalta kun ei ole mitään järjestelmää niin sitten se on aika siitä henkilöstä kiinni, että kuka jaksaa.” (Tek.J.1)*

*”On ollut tietynlailla yhtenäinen järjestely, mutta se, miten [dataa] on täydennetty, on kyllä ihan tekijänsä näköistä. Siinä on selkeästi näkynyt, että mikä vastuualue on kenelläkin ja... miten kokee sen mitä datasta täytyy saada irti.” (Tek.J.2)*

Manuaalinen datan käsittelyprosessi ja heikko turvallisuus vaikuttivat jo datan uskottavuuteen. Järjestelmien puute johti selvästi siihen, että pelko inhimillisistä virheistä oli merkittävä tekijä dataa käsiteltäessä ja käytettäessä.

*”Mitä haasteita siinä [vanhassa järjestelmässä] on: se on manuaalisesti ylläpidettävä, ei ole välttämättä ajan tasalla, on ehkä vähän sinnepäin. Sitten tavallaan se luotettavuus, koska se on kaikki manuaalista.” (TalousP.)*

Sekä Excel-taulukossa että CAD-tiedostoissa tiedostettiin olevan omat ongelmansa. Excel-puolella koko tuotehallinta oli yhdessä tiedostossa, jota pystyi samaan aikaan muokkaamaan vain yksi käyttäjä. Molemmissa oli myös hyvin rajalliset mahdollisuudet rajoittaa käyttäjien vapauksia tehdä muutoksia. Turvallisuuden osatekijän tietojärjestelmänäkökulma ei näin ollen toteutunut kunnolla. Useamman samanaikaisen muokkaajan ongelman voidaan katsoa vaikuttaneen negatiivisesti myös saatavuuden käyttäjänäkökulmaan eli helppokäyttöisyyteen. Lisäksi puutteita havaittiin luotettavuuden tietojärjestelmänäkökulmassa, sillä kummassakaan (Excel ja CAD) ei säilynyt historiatietoa.

*”... rupesi meneen todella hankalaksi se [Excelin] käyttö, kun yks vaan kerrallaan voi käyttää ja kuitenkin olisi monta käyttäjää ja monen tarvis tehdä sinne asioita.” (Suunnittelija)*

*”Jos toinen oli tallentanut muutoksen ja toinen oli ennen sitä muutoksen tallentamista avannut sen saman Excelin, niin se ei tavallaan osannut tehdä niitä kummankin muutoksia siihen samaan Exceliin vaan se loi sitten uuden Excelin. Jos sitä ei huomannut niin sitten ne tehdyt muutokset ei tullut oikeasti alkuperäiseen Exceliin.” (HankintaV.)*

*”Ehkä suurimpana haasteena siinä on se multimuokkaus, monen käyttäjän yhtäaikainen käyttö on ihan selkee puute. Ja sitten se, että se on niin avoin käyttäjän tekemille muutoksille. Ja siinä ei ole sitä historian säilytystä, että näkisi niitä tapahtuneita muutoksia.” (TuoteP.)*

CAD-tiedostojen hallinta oli täysin oma maailmansa, mikä viittasi jatkuvuuden ongelmaan. Puutteita havaittiin jälleen myös turvallisuuden tietojärjestelmänäkökulmassa, koska käyttäjien oikeuksia ei ollut juurikaan pystytty rajoittamaan.

*”CAD-maailman nimikehallinta... oli periaatteessa aika paljon sekavampaa, koska sitä tehtiin CAD:ssa ja siellä oli enemmän käyttäjävapauksia” (TuoteP.)*

*”... se data siinä Excelissä on varmaan ollut ihan suhteellisen hyvälaatuista, mutta se ei ole ollut se koko totuus, koska osa merkittävästä datasta on ollutkin sitten siellä SolidWorksin puolella, missä dataa on hallittu ihan eri tavalla.” (Tietojärj.P.)*

Koska yksittäiset tiedostot, kuten CAD-tiedostot, piirustukset ja dokumentit, olivat tallennettuina perinteiseen kansiorakenteeseen, datan saatavuus muodostui monesti ongelmaksi. Tiedostojen löytäminen oli paikoin oman tiedon varassa, joten saatavuuden tietojärjestelmänäkökulman lisäksi myös käyttäjänäkökulmassa oli tässä suhteessa puutteita.

*”Voi olla, että tehdään paljon päällekkäistä työtä. Joku on jo tehnyt jonkun dokumentin, kukaan ei vaan tiedä kun sitä ei voi hakea oikeen millään. Jos et tiedä just exact nimeä tai tiedä sitä kaveria, että vois käydä kysymässä, niin sitten et oikeestaan löydä sellasta dokumenttia, joka on joskus ehkä tehty.” (Tek.J.1)*

*”Jos et tiedä missä [tiedosto] on, niin aika haasteellisesti löydettävissä kun on vähän mitään indeksointia, hakutoimintoo tai versioylläpitoa. ... Kyllähän se [data] löytyy, mutta kyllähän se paljon hiljaisen tiedon takana on.” (TalousP.)*

*”Ainut mikä siinä [piirustusten löytämisessä kansiorakenteesta] auttaa on se, että tietää missä ne piirustukset oikeesti on. Siinä pitää aika paljon osata ulkoa, että sieltä löytää.” (HankintaV.)*

Jotkut dataa hallinnoivista haastateltavista kokivat myös erityistä huolta datan turvallisuudesta, sillä tiedostojen poistaminen onnistui kaikilta käyttäjiltä. Kansiorakenteen todettiin aiheuttavan myös sen, että datan varmuuskopioiminen oli todella hankalaa, mikä aiheutti riskin turvallisuuden prosessinäkökulmasta.

*”Sitten äkkiä voi olla myöskin, että tieto ei ehkä pysy tallessakaan. Jos jollain käy mielessä, että enpä tarvi noita vanhoja revisioita, painan deletee, nehan katoavat normaalista kansiojärjestelmästä, eihän niitä sitten löydy enää mistään.” (Tek.J.1)*

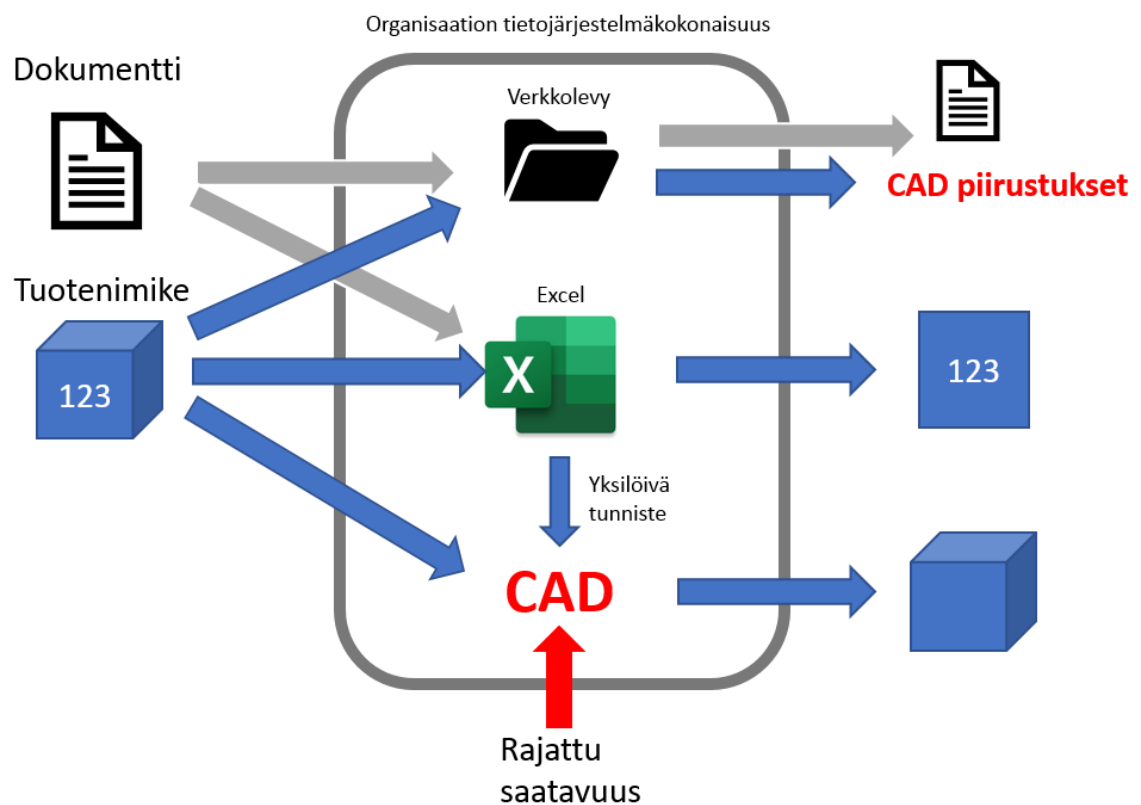
*”[Vanhassa järjestelmässä] tulee helposti [tiedostotasolla] korruptoituneita versioita tai sitten päällekkäisiä tallennuksia tai samaa nimikenumeroa käytetään useaan kertaan.” (Tek.J.2)*

Myös luotettavuuden prosessinäkökulmassa oli parantamisen varaa, sillä loppukäyttäjien mukaan data ei aina ollut ajan tasalla. Tämä johtui monesti siitä, että Excelissä ja CAD-mallissa oli omat rakenteensa. Toiseen tehty muutos piti päivittää manuaalisesti myös toiseen. Tätä puolestaan vaikeutti se, että CAD-malleihin oli erittäin rajallinen pääsy. Lisäksi jotkin osat eivät esiintyneet CAD-mallissa ja jotkin puolestaan eivät esiintyneet Excelissä.

*”No lähinnä ne [virheet] tuli juuri sieltä rakenteesta ja siitä että rakenne ei ollut välttämättä aina päivittynyt viimeisimmälle tasolle, revisio-ongelmia osien suhteen ja data saattaa olla vanhaa ja joitain yksittäisiä osia saattaa sieltä jopa puuttua. Jotkin asiat on vaan tiedon varassa.” (VarastoV.)*



Datassa ja datan käytössä on siis ennen tietojärjestelmä uudistusta ollut paljon erilaisia ongelmia. Positiivinen asia oli se, että haastateltavat tiedostivat ongelmat varsin hyvin. Kun ongelma on tiedostettu, sitä voidaan alkaa korjaamaan. Kuvassa 6.1 on havainnollistettu kohdeyrityksen tuotetietojen luominen, säilytys ja käyttö karkealla tasolla lähtötilanteessa.



**Kuva 6.1.** Kohdeyrityksen lähtötilanne tuotetietojen luomisen, säilytyksen ja käytön suhteen.

Tuotenimike siis tallennettiin uutena rivinä tuotehallinnan Exceliin, josta sille saatiin yksilöivä tunniste. Tunniste kopioitiin käsin CAD-ohjelmistoon, jossa nimikkeelle luotiin CAD-malli. Malli saattoi sisältää tietoa nimikkeestä, jota ei tallennettu Exceliin. CAD-mallista luotiin pdf-piirustus, joka tallennettiin vielä verkkolevylle. Näin ollen yhden tuotenimikkeen dataa oli kolmessa paikassa, jotka eivät vaihtaneet tietoa keskenään. CAD on nostettu omaksi "sijainnikseen" kuvaan, sillä CAD-mallien tietoihin pääsi käsiksi vain CAD-ohjelmiston kautta, ja näitä ohjelmistoja oli vain CAD-suunnittelijoilla. On myös huo-

mattava, että CAD-ohjelmistossa tehdyt muutokset piirustuksiin eivät päivittyneet automaattisesti verkkolevyllä oleviin pdf-piirustuksiin, vaan tämäkin päivitys oli käyttäjän vastuulla.

Keskeinen huomio kohdeyrityksen datan hallinnasta lähtötilanteessa oli myös se, että tuotehallinnan Excel ei suinkaan ollut ainoa Excel. Tuotehallinnan lisäksi kullakin toiminnolla, kuten varastolla ja hankinnalla, oli oma Excelinsä, jossa käytännössä samaa dataa hallinnoitiin omalla tavallaan ja erillään muista Exceleistä. Tämä aiheutti merkittävää datan epäjatkuvuutta ja virheitä.

Excelistä saatava tieto koettiin tulkinnanvaraiseksi, joten nimikkeen hahmottaminen vaati usein vähintään kahden sijainnin hyödyntämistä, usein Excelin ja piirustusten. Myös dokumentit tallennettiin sekä Exceliin uutena rivinä että verkkolevyille. Kohdeyrityksen lähtötilanteen summaa hyvin tietojärjestelmäpäällikön sitaatti:

*”Loppupeleissä se [datan laatu] ei ole kuitenkaan ollut kauhean hyvälaatuista.”  
(Tietojärj.P.)*

### 6.3 Datan laatu nykyhetkellä

Tietojärjestelmien käyttöönotolla nähtiin heti olevan positiivisia vaikutuksia datan laatuun. Mielenkiintoisena huomiona osa haastateltavista näki vanhan tuotehallinnan järjestelyn olleen hyvin kevyt ja sopeutuva, kun taas toiset näkivät sen olleen erityisen kankea. Tarkkuuden tietojärjestelmänäkökulman toteutuminen koettiin siis eri tavoin.

*”Siinä mielessä se [Excel] on ollut hyvä, että alkuun oli sellainen kevyempi versio ja kun meidän työ on kuitenkin ollut hyvin moninaista ja vähän joka suuntaan sinkoilua niin Excel on pysynyt siinä vauhdissa sitten mukana. Mutta nyt on hyvä, kun toiminta on kasvanut niin... järjestelmäasiat on edennyt siinä mukana ja tehneet toiminnasta vähän järjestelmällisempää.” (HankintaV.)*

*”Meidän vanha järjestelmä (Excel) on ollut tosi kankea. Sieltä on ollut vaikeaa löytää ylipäättään mitään. Näistä uusista [PLM & ERP] on odotettavissa, että käyttö on huomattavasti sujuvampaa. Vanhassa järjestelmässä oli niin villin län-*

*nen tekniikalla vaan heitetty ne kansiorakenteetkin ja muut; se oli aivan liian monimutkainen. Plus ne [tiedosto]polut alkoi olemaan niin pitkiä, ettei [tiedostot] enää auennut.” (VarastoV.)*

Excel-datan ja CAD-datan yhdistäminen yhteen PLM-järjestelmään nähtiin yhtenä tietojärjestelmä uudistuksen suurimmista haasteista. Tämä kuitenkin paransi huomattavasti datan jatkuvuutta. Haasteita aiheutti se, että jatkossa yhden tietomallin oli kyettävä kattamaan sekä Excel- että CAD-datan tarpeet.

*”Se, mikä siinä on ehkä tullut haasteeksi, että kun CAD-maailma on ollut omillaan ja sitten muu tuoterakenne on ollut omillaan, niin niille ollaan saatu eriytettyä toisistaan riippumattomat tärkeät arvot. Nyt kun ne ajetaan samaan järjestelmään niiden pitää olla myös samassa formaatissa.” (TuoteP.)*

Excel-datan ja CAD-datan yhdistämisessä erityiseksi haasteeksi muodostui tuoterakenne. Tuoterakenne jouduttiin tietojärjestelmä uudistuksen yhteydessä laajalti päivittämään paitsi Excel-datan ja CAD-datan yhdistämisen vuoksi, myös tulevaa ERP-järjestelmää ja tuotantoa varten.

*”Nyt me vasta saadaan se läpinäkyvyys meidän täydelliseen tuoterakenteeseen. Se on yhdistelmä siitä CAD-mallista, joka sisältää Excelistä puuttuvat osat, ja Excelistä lisättynä ne osat, joita ei ole CAD-mallissa. ... [Excelin ja CAD:n yhdistäminen] paljastaa niitä puutteita, mutta myöskin se valmistaa meitä siihen, että kun tulevaisuudessa meidän pitää oikeasti tietää joka ikinen osa mitä meidän koneessa on, niin tällä me saadaan se läpinäkyvyys.” (Tietojärj.P.)*

*”Yksi iso haaste on se, että tuoterakenne [PLM:ssä] muodostuu nyt lähtökohtaisesti CAD-mallin pohjalta. Jos se CAD-mallin puurakenne ei ole paras mahdollinen niin se tarkoittaa, että tuoterakenne PLM:ssä ei ole paras mahdollinen... ja sen korjaaminen on aika iso työ.” (TuoteP.)*

Excel- ja CAD-datan sekä dokumenttien yhdistäminen yhteen tietojärjestelmään oli kuitenkin tärkeä askel kehityksessä ja siitä koettiin välittömästi hyötyjä. Muun muassa datan

saatavuus parani, sillä kaiken datan, erityisesti teknisen datan, löytymisen yhdestä paikasta katsottiin helpottavan valtavasti tiedon hakemista.

*”Haaste on ollut se, että vaikka olet ennen löytänyt sen nimikkeen, niin et välttämättä ole löytänyt sitä dataa.” (Tek.J.2)*

Tietojärjestelmien mahdollistama tietomallien, tietokenttien, käytäntöjen ja prosessien (esimerkiksi muutosprosessit) standardoiminen ja muutoshistoria nähtiin suurina etuina aiempaan käytäntöön verrattuna. Nämä tulevat parantamaan merkittävästi kohdeyrityksen datan laatua jatkuvuuden, luotettavuuden sekä turvallisuuden kannalta.

*”Isoimpana [hyötynä] on se muutoshistoria ja hallitut muutokset. ... [Lisäksi] käyttäjän syöttämää dataa on vähennetty ja... data syötetään esivalittuihin listoihin. Se tulee parantamaan [datan laatua].” (TuoteP.)*

*”Nimikkeet voidaan jäädyttää. Kun ne on hyväksytty-tilassa... niitä ei kukaan pääse vahingossa muuttaleen ja jos jotain muutetaan niin siitä jää aina jälki järjestelmään; koska ja mitä on muutettu ja kuka [muutoksen] on tehnyt. Historiatieto pysyy paremmin tallella.” (Suunnittelija)*

*”[Muutos- ja hyväksyntä]prosessit on kaikki standardoituja, ne menee kaikki sen saman kaavan mukaan. Kellään ei ole mitään kysymyksiä, että mites tässä pitäisi toimia.” (Tietojärj.P.)*

PLM-järjestelmän katsottiin parantavan huomattavasti datan yhteiskäyttöä, joka oli keskeinen ongelma vanhassa järjestelyssä. Käyttäjien näkökulmasta myös datan saatavuus parani PLM-järjestelmän myötä merkittävästi.

*”Onhan se hienoa, että on yksi selkeä tapa, miten me toimitaan ja mistä data löytyy, missä on viimeisimmät revisiot ja kaikki tämmöset.” (Tietojärj.P.)*

*”... [dokumenttien] datan laadussa en näe kovinkaan isoa muutosta, mutta sen hallittavuus on parantunut merkittävästi. [Dokumentteja] ei joudu enää etsimään niin vaikeasti vaan ne on helpompaa löytää.” (TuoteP.)*

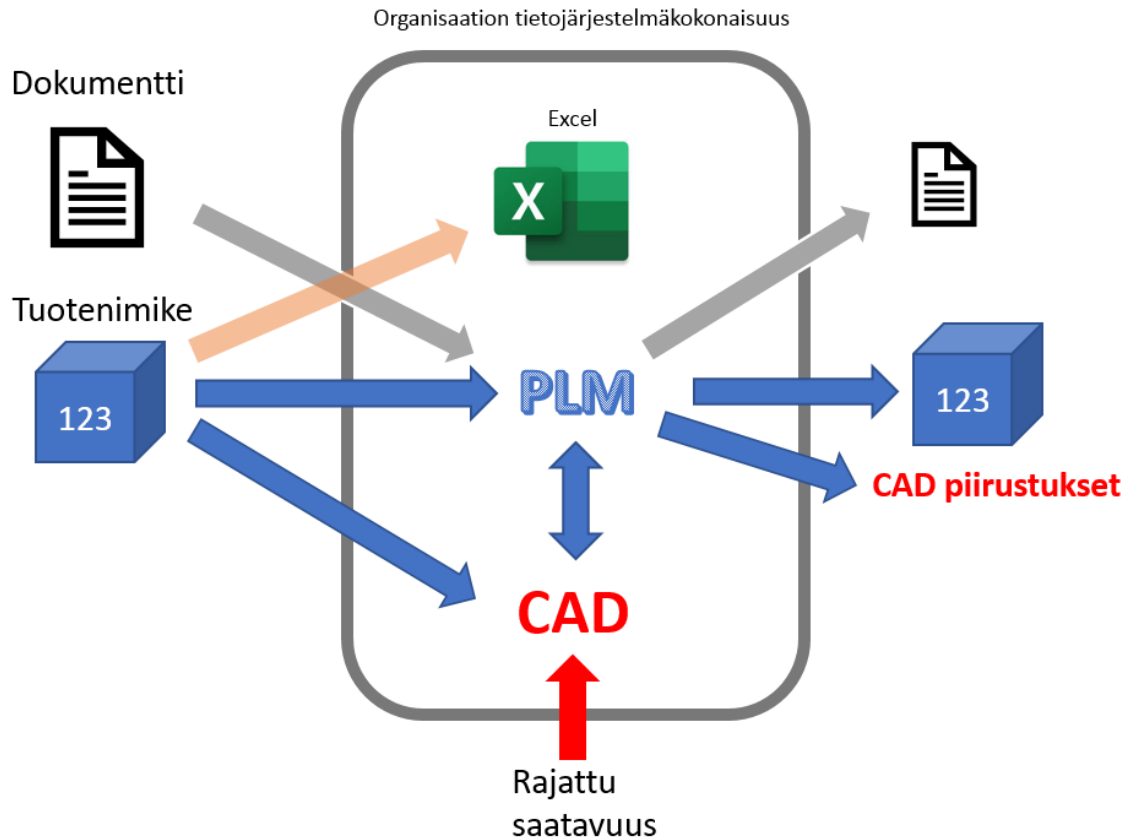
Datan hallinnoijat kiinnittivät huomiota myös datan turvallisuuden paranemiseen PLM-järjestelmän ansiosta erityisesti data- ja tietojärjestelmänäkökulmista. Järjestelmällä pystytään ”jäädyttämään” data tiettyyn pisteeseen, jonka jälkeen kaikista muutoksista jää jälki. Järjestelmän mahdollistama käyttöoikeuksien hallinta myös vähensi valtavasti riskiä datan hävittämisestä vahingossa.

*”Nyt kun me ollaan saatu [komponentteja] hyväksyttyä, ei ole myöskään liian isoja riskejä siinä, että joku pystyisi rikkomaan tai deletoimaan jotain. ... Plus meidän järjestelmästä otetaan koko ajan back uppeja ja on mietitty myös sitä, miten me varmennetaan että meidän data säilyy turvassa. Se on yks jopa isoimmista hyödyistä tässä kohtaa.” (Tietojärj.P.)*

Tiivistettynä PLM-järjestelmä otettiin kohdeyrityksessä mieluusti vastaan. Toisaalta järjestelmän käyttöönotossa on ollut myös selviä hankaluuksia. Datan jatkuvuus muutosvaiheessa nähtiin yhtenä näistä hankaluuksista, koska vanhan järjestelmän käyttöä ei pystytty lopettamaan välittömästi uuden järjestelmän käyttöönoton jälkeen, vaan näitä jouduttiin käyttämään rinnakkain.

*”Se on tavallaan myöskin pieni kompastuskivi sen suhteen, että eihän sitä järjestelmää tule käytettyä ennen kuin se on täydellinen ja se ei ole täydellinen ennen kuin sitä on käytetty, tai lähellä täydellistä.” (VarastoV.)*

Kuvassa 6.2 on havainnollistettu tuotetiedon käsittelyn nykyhetkeä. Kuvassa PLM-järjestelmä on korvannut verkkolevyn sekä osittain tuotehallinnan Excelin.



**Kuva 6.2.** Nykyhetken tilanne kohdeyrityksen tietojärjestelmien suhteen.

Kuvasta havaitaan, että tilanne on selkiytynyt aiemmasta huomattavasti. Lähtökohtaisesti kaikki data saadaan ulos PLM-järjestelmästä. CAD-ohjelmisto on yhä rajattu, ulkoinen kokonaisuus, mutta se siirtää automaattisesti tietoa PLM-järjestelmän kanssa, joten manuaaliselta työltä (ja virheiltä) vältetään tässä yhteydessä. Lisäksi PLM-järjestelmä mahdollistaa CAD-mallien esikatselun, mikä helpottaa huomattavasti nimikkeiden hahmottamista. Tuotehallinnan Excel on yhä taustalla mukana turvaamassa datan säilyvyyttä mahdollisissa ongelmatilanteissa.

## 6.4 Datan laatu tulevaisuudessa

Haastateltavilla oli melko yhtenevät näkemykset kohdeyrityksen datan tilanteesta tulevaisuudessa. Kaikki data tulee löytymään yhdestä paikasta samassa formaatissa. Tietojärjestelmän ohjaavat datan laadukkaampaan suuntaan tarjoamalla selkeät tietomallit ja pohjat täytettäväksi.

*”Datan laadun pitäisi olla parempaa sikäli, että on pakolliset kentät ja myös se, mikä tuli tässä käyttöön, että nimitykset pakotetaan listasta. Se, että ne on yhdenmukaisia, helpottaa asioiden löytämistä.” (Suunnittelija)*

*”Jatkossa on vielä helpompi siivota sitä datamuotoa, esim. standardiruuvit ja -mutterit pystytään päivittämään siihen, että ne kaikki on tismalleen samassa muodossa.” (TuoteP.)*

Selkeät tietomallit eli panostaminen tarkkuuden datanäkökulmaan sekä datan löytymisen yhdestä järjestelmästä tulevat parantamaan datan saatavuutta todella merkittävästi. Keskitetyn datan hallinnan katsottiin myös vähentävän virheiden mahdollisuuksia ja ylimääräistä työtä, joka nousi esiin haastatteluissa useasti. Selvästi yksi odotettu ominaisuus tietojärjestelmältä oli mahdollisuus tallentaa enemmän dataa yhden nimikkeen alle, mikä parantaa datan tarkkuutta ja saatavuutta.

*”Tietojärjestelmä uudistuksen myötä data pitää saada keskitettyä vielä paremmin yhteen paikkaan, että copy-paste järjestelmästä toiseen tai asioiden tuplakirjauksia eri järjestelmiin [ei tarvita]. Se on niin hukkaan heitettyä aikaa, kun vaan voi olla.” (TuoteP.)*

*”Löydetään mallit ja löydetään ne suunnittelukriteeristöt, millä vanhat komponentit on suunniteltu ja sen jälkeen pystytään periaatteessa suunnittelemaan se sama komponentti uudestaan vanhaan tuotteeseen.” (Tek.J.2)*

Hyvälaatuisen datan katsotaan tulevaisuudessa tukevan kohdeyrityksen liiketoimintaa hyvin monen tekijän kannalta. Esimerkiksi tarkkuuden, saatavuuden ja jatkuvuuden prosessinäkökulmat tulevat paranemaan. Tämän katsottiin edelleen pienentävän yrityksen kustannuksia ja parantavan tuotteiden ja palvelun laatua.

*”Hyvälaatuinen data pienentää kustannuksia hankinnan näkökulmasta, valmistuksessa tulee vähemmän virheitä ja meidän koneet on paremman laatuista. Niissä ei esiinny mitään laatu poikkeamia tai laatu virheitä, johtuen vaikka siitä, että meidän tuoterakenteet ei olisi ajan tasalla. ... Kun myynti menee asiakastapaamisiin,*

*niin he ovat uskottavampia, koska he löytävät sen datan mitä he tarvitsevat siihen myyntitapahtumaan. [Tarvittava data] löytyy selkeämmin sen hyvälaatuisen datan kautta, siihen pystyy luottamaan ja he pystyy muodostamaan myynticasensa hyvälaatuisen datan pohjalle” (Tietojärj.P.)*

Turvallisuuden tietojärjestelmänäkökulma tulee paranemaan, sillä dataan kohdistuva muutosprosessi tietojärjestelmäkokonaisuudessa on selkeä, joustava ja standardoitu prosessi. Lisäksi kaikista muutoksista tallentuu järjestelmään historiatieto, joten myös datan seurattavuus paranee. Kaikella datalla ja toiminnalla on yksi selkeä pääjärjestelmä.

*”Taustalle ei tarvitse niin massiivista järjestelmää. Se kuitenkin korostuu, että on joku paikka, jossa datan masteruus (hallinnointi) on. Se on tärkeää.” (Tietojärj.P.)*

Haastatteluvaiheessa oli nähtävissä paljon odotuksia myös tulevalle ERP-järjestelmälle. Tähän asti kohdeyrityksellä ei ole ollut mitään järjestelmää valmistuksen, tilausten ja toimitusten hallintaan. Tilauksia ja toimituksia on hallinnoitu Exceleissä, kun taas valmistus on ollut lähinnä hiljaisen tiedon varassa.

ERP-järjestelmä tulee integraation kautta käyttämään PLM-järjestelmän tuotedataa. Yksi tulevaisuuden vaatimus tietojärjestelmäkokonaisuudelle on selkeä vastuunjako, mitä tehdään missäkin järjestelmässä. Tämä on erityisen tärkeä asia jatkuvuuden tietojärjestelmänäkökulmasta katsottuna.

*”[ERP+PLM-integraation] pitäisi mahdollistaa datan suoraviivaista käsittelyä ja automatisoida sitä järjestelmästä toiseen siirtymistä ja ennen kaikkea tukea sitä, että, mitä PLM puolella tarvitaan, niin se löytyisi PLM:stä ja sitten taas mitä ERP:ssä tarvitaan. Oleelliset datat vaihtaa integraation myötä tietonsa, mutta se rajataan selvästi, että jos joku asia tehdään ERP:n puolella niin se tehdään ERP:n puolella eikä siihen tarvitse ottaa kantaa PLM:ssä.” (TuoteP.)*

Toisaalta, jos jatkuvuuden tietojärjestelmänäkökulmassa alkaa ilmenemään puutteita, tämä aiheuttaa merkittäviä ongelmia. Tietojärjestelmiin liittyvänä riskinä nähtiinkin järjestelmien ajautuminen erilleen.



*”PLM on tuotetiedon hallintajärjestelmä ja ERP vaan peilaa sitä. Ettei valuttaisi siihen, että meillä on kaks eri järjestelmää, jotka molemmat elää omaa elämänsä.” (TalousP.)*

Hyvälaatuisen datan ansiosta tietojärjestelmäkokonaisuus tulee mahdollistamaan nopean reagoinnin liiketoiminnan tarpeisiin ja avaamaan täysin uusia mahdollisuuksia mm. elinkaaren hallinnan ja verkkopalveluiden suhteen. Toisaalta yhden datan käyttäminen moneen tarkoitukseen asettaa omat vaatimuksensa mm. tietomallin ja jatkuvuuden suhteen.

*”Sen [PLM+ERP] pitää mahdollistaa se, että me pystytään ensinnäkin hyvin nopeasti vastaamaan asiakastarpeisiin [tilaustoimitusketjun] alkupäässä, pystytään määrittelemään mikä yksilö me halutaan asiakkaalle toimittaa, pystytään valmistamaan ja hankkimaan siihen liittyvät osat helposti ja sujuvasti osana tilaustoimitusketjua ja sitten myös sen jälkeenkin, kun kone on toimitettu asiakkaalle, meillä on selkeä tieto siitä mitä me ollaan toimitettu, mihin ja kenelle. Kun se kone myös elää sitten elinkaaren aikana, niin me pystytään ylläpitämään sitä kokonaista tuoterakennetta ja me tiedetään tasan tarkkaan mitä siellä asiakkaalla on missäkin sen elinkaaren vaiheessa. Me tarjotaan siinä asiakkaalle sitä lisäarvoa. Me pystytään palvelemaan asiakasta paljon paremmin, kun me tiedetään siitä laitteesta kaikki ja pystytään silloin tukemaan tosi nopeasti asiakasta after salesin osalta.” (Tietojärj.P.)*

*”Kun tuotteita huolletaan ja niille tehdään mahdollisesti jotain pieniä muutoksia elinkaaren aikana, niin me nähdään selkeästi ne koneyksilöt ja pystytään tietämään, jos siihen koneeseen on tehty vaikka viiden vuoden kohdalla joku huolto. Se elinkaari olisi silloin näkyvissä. ... [Lisäksi] pystytään luomaan tällaisia uusia verkkokauppahenkisiä palveluita, mistä voidaan osat helposti myydä suoraan ilman, että me formuloidaan dataa siinä välissä tai että luodaan jotain [erillistä] varaosajärjestelmää tai varaosatietokantaa vaan että se tulee ihan samasta datasta. Data ajetaan vaan eri muotoon asiakkaalle nähtäväksi. Silloinhan voidaan jo sanoa, että data on aika hyvää, jos se siitä, mitä se meillä on, kääntyy suoraan asiakkaalle näytettävään muotoon.” (Tek.J.2)*

Näin suuri tietojärjestelmä uudistus on kuitenkin altis riskeille. Yleisimmin esiin noussut riski oli riippuvuus tietojärjestelmästä eli turvallisuuden prosessinäkökulma. Mitä laajemmin tietojärjestelmä on yrityksessä käytössä, sitä laajemmin myös siinä esiintyvät ongelmat vaikuttavat yrityksessä. Toinen huomio oli järjestelmän mahdollisesti aiheuttama riippuvuus järjestelmän tarjoajasta tai kolmannesta osapuolesta.

*”Näkisin, että järjestelmäriski on se suurin... Järjestelmä ei aiheuta datalla niinkään riskiä. Järjestelmä voi mennä rikki. Ja se piste, kun järjestelmä ei enää napin painalluksesta tee sitä mitä se teki eilen, niin mitä siinä pisteessä tehdään... Se voi pahimmassa tapauksessa tarkoittaa, että silloin ei tapahdu yhtään mitään, koska se kaikki data on siinä järjestelmässä. Jos järjestelmästä ei siirry dataa tai sieltä ei saa pihalle dataa, niin pahimmassa tapauksessa tarkoittaa, että ei voida toimia.” (TuoteP.)*

*”... yksi riski voi tietenkin olla, että kuinka järjestelmäosaamista on itsellä. Jos tulee vaikea kysymys, ei tunneta sitä omaa järjestelmää, ei osata tehdä sille mitään, esim. koodata jotain asiaa sinne, niin sittenhän se on hankalaa. Sitten otetaan yhteyttä johonkin palveluntarjoajaan, joka taas ilmoittaa, että no ei mekään tätä osata, se on tuolla ulkomailla se erikoisosaaminen. Sitten ollaankin jossain Amerikassa seuraavaksi ja sillä kaverilla on just 200 pyyntöä ja me ollaan rivissä viimeinen ja sitten se on parin kolmen kuukauden päästä tehnyt sen muutoksen meille.” (Tek.J.1)*

Haastateltavia huolestutti myös tietojärjestelmäkokonaisuuden integraatio. Aiemmassa järjestelyssä yksittäinen virhe jäi hyvin paikalliseksi. Uudessa integroidussa kokonaisuudessa aikaisessa vaiheessa tehty virhe siirtyi automaattisesti kaikkialle.

*”Jos PLM:n data dyykkaa, niin sen jälkeen ongelmat siirtyy ERP:n puolelle. ... Monesti ihmiset käyttää ERP:stä jotain tiettyä osiota ja tiettyä toiminnallisuutta. He ei välttämättä ymmärrä sitä koko polkua, että kun mä tässä toimin näin, niin mihin se mahdollisesti vaikuttaa... Sen takia ERP:n toiminnallisuudet ja toimintatavat ja meidän toimintaprosessit pitäisi rakentaa niin, että me kokoajan mietitään sitä, että tuleeko tälläin toimien varmasti hyvälaatuista dataa pihalle” (TalousP.)*

*”... inhimillinen virhe aiheuttaa mahdollisesti myös enemmän ongelmia, koska se toteutuu sitten useampaan kappale määrään. ... Siihen pitää pystyä luottaan silloin, että tieto on oikeeta.” (VarastoV.)*

*”[PLM] on kuitenkin pohja ERP:n toiminnalle... Varmistettaisiin, että se PLM:än tuotu data on oikeasti kunnossa ja ne rakenteet on ajan tasalla ja ajan mukaisia, jotta se virheellinen tieto ei tulisi sitten ERP:hen saakka, koska sitten se joudutaan sieltä taas tuomaan takaisin ja joudutaan tulemaan takaisin PLM:än, jossa se data täytyy korjata... Jos meillä on ollut virheellistä dataa, eikä huomata sitä virhettä, niin se virhe menee eteenpäin ja eteenpäin tilaustoimitusketjussa ja pahimmillaan tänne meille tulee virheellinen osa, osia tulee väärät määrät tai virheellinen osa menee asiakkaalle” (HankintaV.)*

Toisaalta toimiva integraatio myös saa aikaan sen, että jos virhe löydetään, sen tiedetään olevan kaikkialla. Tämä edesauttaa osaltaan virheiden leviämistä, mutta toisaalta tällöin myös yhdellä korjauksella kyseinen virhe saadaan korjattua kaikkialta sen sijaan, että virhettä jouduttaisiin etsimään useasta sijainnista. Lisäksi tällaisten virheiden katsottiin olevan huomattavissa helpommin kuin vain yhdessä paikassa olevien virheiden.

*”Esimerkiksi, kun revisiotieto katsotaan vaan yhdestä paikasta, niin jos se menee väärin, sitten tiedetään, että se menee kaikkiin väärin ja silloin siihen tulee kiinnitettyä helpommin myös huomiota. Todennäköisesti tieto tulee olemaan ajantasaisempaa helpommin ja se on kaikilla löydettävissä helposti. Kun tosiaan on ollut monta järjestelmää, mistä on haettu tietoja, niin siinä tulee ongelmia, mutta me tullaan pääseen niistä tällä ylitse.” (VarastoV.)*

Lähtökohtaisesti järjestelmästä toiseen ei siirretä ylimääräistä dataa, koska se hankaloittaa datan saatavuutta (ylimääräistä dataa). On tärkeää huomata, että ERP:ssä on erilaiset datatarpeet kuin PLM-järjestelmässä. Tämä asettaa datalle omat haasteensa ja vaatimuksensa, koska samaa dataa on pystyttävä kuitenkin käyttämään useammassa tietojärjestelmässä ja eri tarkoituksiin.

*”Siirretään ainoastaan oleellinen data järjestelmästä toiseen.” (TuoteP.)*

*”Sitä [dataa] on pystyttävä käyttämään monessa järjestelmässä” (Tietojärj.P.)*

Joissakin haastatteluissa nousi esiin huoli siitä, että tietojärjestelmiin aletaan luottaa liikaa. Tietojärjestelmien saatetaan itsessään uskoa korjaavan datassa olevat laatuongelmat, jolloin on riskinä, että sopiva kriittisyys ja virheiden mahdollisuus unohdetaan. Tämä osoittaa sen, että datan maine ja uskottavuus (luotettavuuden käyttäjänäkökulma) voi kääntyä negatiiviseksi paitsi heikossa tapauksessa myös liian hyvässä tapauksessa. Yrityksen ulkopuolelle pääsevät virheet heikentävät yrityksen uskottavuutta ja mainetta.

*”Kyllä se mun mielestä perustuu ihan siihen, että oltaisiin tarkkoja ja kriittisiä sille omalle tekemiselle ja kyseenalaistettaisiin terveellä järjellä sitä, että onko tää nyt oikein vai onko tässä nyt virhe. Tai jos löydetään jostain pienikin virhe niin siihen puututtaisiin rohkeasti, koska vaikka se virhe tuntuu pieneltä itselle siinä omassa toimenkuvassa ja työtehtävässä, niin sama virhe voi jossain muualla aiheuttaa tosi paljon päänsärkyä tai esimerkiksi rahallisia menetyksiä.” (HankintaV.)*

*”[Virheistä voi seurata] että me vaikkapa toimitetaan vääriä osia varaosiksi tai me valmistetaan väärillä osilla tai tehdään jotain, joka huonontaa sen koneen toimivuutta, ja sitä kautta [kohdeyrityksen] uskottavuus alkaa heikkenemään. Tai meillä tulee ongelmia meidän koneiden kanssa tai meidän toimitusajat alkaa yhtäkkiä venymään johtuen siitä, että meillä ei ole oikeita osia tai me ostetaan väärillä kuvilla. ... Ja sitten tietenkin ylipäättään sellainen kaaos täällä yrityksen sisällä, että esimerkiksi meidän talouspuoli... tai varastonhallinta ei enää tiedä, mitä komponentteja meillä on oikeasti varastossa. Se voi nostaa meidän varastoarvoa ihan valtavasti. Se on iso uhka ja se myös työllistää paljon ihmisiä, jos näin pääsee tapahtumaan. Ja myös se, että ei pystytä raportoimaan. Johto ei pysty tekemään oikeita päätöksiä, koska meidän luvut ovat vääristyneitä johtuen huonolaatuisesta datasta.” (Tietojärj.P.)*

Datan laadun kehittäminen ja ylläpitäminen vaativat myös jatkuvia toimenpiteitä. Datanäkökulmasta tämä edellyttää, että tietomallit ovat oikein ja ajan tasalla. Tietojärjestelmien ja tietomallien on myös pystyttävä reagoimaan muuttuviin vaatimuksiin.

*”Ensinnäkin se, että tietomallien on oltava kunnossa. Niiden on oltava mietitty loppuun asti.... Meidän pitää pystyä muuttumaan tosi nopealla tahdilla ja tietomallien pitää myös pystyä vastaamaan siihen.” (Tietojärj.P.)*

*”Se pitää olla hyvin mukautuvaa siihen, että mennään muuttuvassa markkinakentässä ja muuttuvien määräysten (standardien) vallassa.” (Tek.J.2)*

Tietomallissa on syytä huomioida myös turvallisuuden datanäkökulma eli tietomallin kestävyys. Tämä viittaa siihen, että data ei ole altis virheille. Muun muassa valmiiden valintalistojen vapaiden tekstikenttien sijaan katsottiin edistävän datan laatua tässä suhteessa.

*”Meillä ei ole paljoa niitä vapaita kenttiä. Jos on, niitä ei hirveästi kuitenkaan hyödynnetä missään. Tavallaan se [tietomalli] on hyvin standardoitua, mutta me pystytään kuitenkin sitä mukauttamaan ja taivuttamaan moneen siltikin. Me ei tiedetä mitkä ne [tulevaisuuden] vaatimukset on, mutta sen datan on muokkautettava tarpeitten mukaan” (Tietojärj.P.)*

*”Kun siellä [ERP:ssä] me conffataan niitä toimintatapoja, miten me ERP:tä käytetään, jos siellä annetaan liikaa vaihtoehtoja tai liikaa vapauksia tai liikaa tilaa tälle möselle human errorille, niin datan laatu huononee helposti.” (TalousP.)*

Vaikka tietojärjestelmät ohjaavat kohti laadukasta dataa ja mahdollistavat paremman datan validoinnin, se ei aina riitä. Virheet lähtevät useimmiten käyttäjistä, joten käyttäjien koulutus ja datan laadun merkityksen esille tuominen ovat tärkeässä roolissa datan laadun kehittämisessä.

*”Laatu lähtee ensimmäisenä käyttäjästä. [Käytön] on oltava riittävän helppoa ja suoraviivaista ja riittävän pieni kynnys siihen toimintaan siirtymiseen niin käyttäjä on motivoitunut ja se näkyy sitten laadussa. ... Sitten suoraan puhdasta laatua, ja*

*nimikelaatua, jos ajatellaan, ehkä se haaste ja hyvyys tulee siitä, että ihmiset näkee sen merkityksen, ... jaksaa nähdä sen vaivan, että tekee sen oikein, eikä oikaise.” (Tek.J.2)*

*”Kaikki koulutetaan hyvin käyttämään järjestelmää ja siihen tekemiseen omistaututaan sillain oikealla tavalla, että jokainen tietää mitenkä sitä käytetään ja mikä on se toimintaperiaate ja esimerkiksi se marssijärjestys, missä vaiheessa tulee uudet revisiot.” (VarastoV.)*

Tietojärjestelmät saatetaan joskus mieltää toimintaa kankeuttavina tekijöinä, erityisesti muutoshallinnan suhteen. Tämä puolestaan lisää järjestelmien vastustusta käyttäjien keskuudessa, minkä voidaan katsoa lisäävät kankeutta entisestään. On kuitenkin syytä pohtia, mistä mahdollinen kankeus johtuu, ja pyrkiä korjaamaan tämä mahdollisimman aikaisessa vaiheessa.

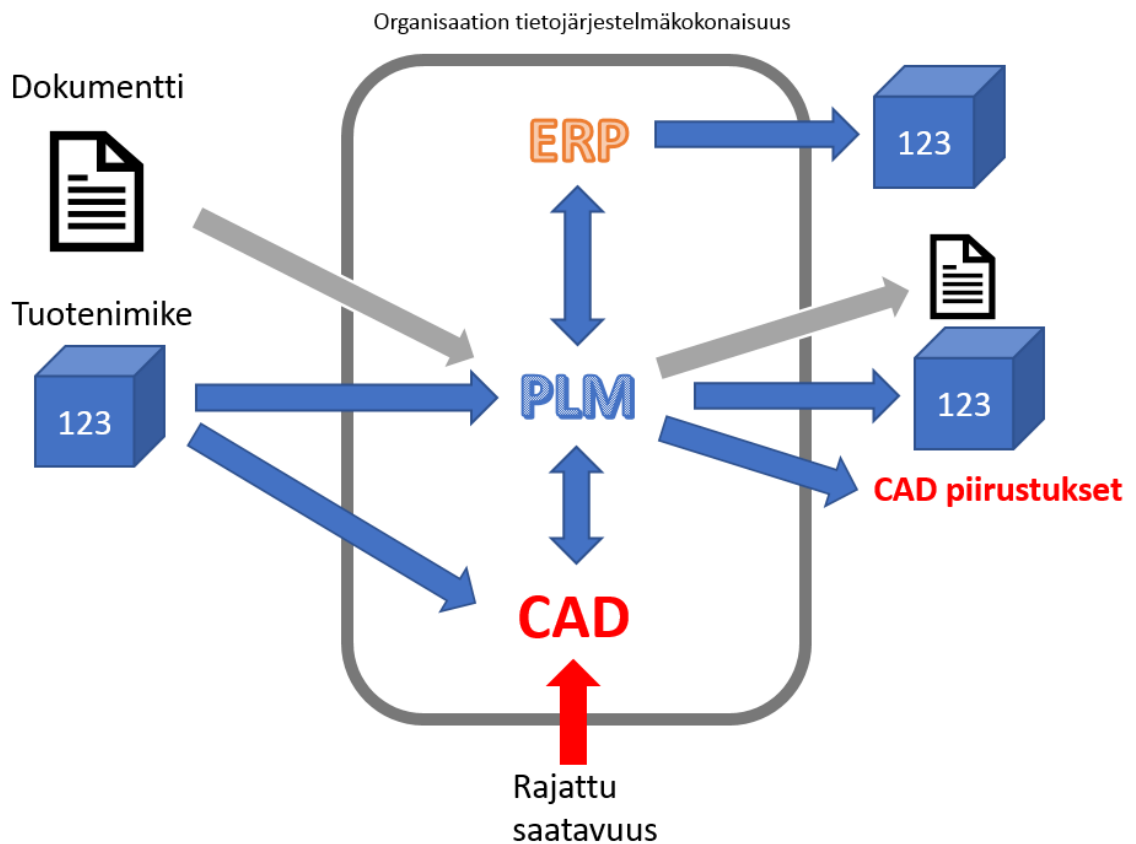
*”Luulen, että se [kankeus] tulee siitä firmasta [eikä tietojärjestelmästä itsestään]” (Tek.J.1)*

Datan laadun hallinta ja ylläpito nähdään niin merkittävänä osa-alueena, että sitä ei voi uskoa täysin pelkästään datan tuottajille ja käyttäjille. Yrityksessä koetaan tarvetta henkilölle, jonka pääsääntöinen tehtävä on huolehtia datan laadusta.

*”Pelkästään se, että käyttäjät tekevät sitä päivittäistä dataa, ei riitä... Nykypäivää on, että meillä pitää olla nimenomaan sellainen rooli, jokainen yritys tarvitsee jonkun, joka vastaa siitä master datasta, jolla on isompi käsitys siitä... mihin sillä [datalla] on vaikutusta ja miten sitä ylläpidetään. Tässä roolissa on jatkuvaa kouluttamista ja sen datan tarkastelua. Missään kohtaa me ei voida pysähtyä vaan ja sanoa, että nyt on hyvä tilanne ja nyt sitten tehdään vaan tälle seuraava vuosi, koska niitä virheitä tulee jatkuvasti ja sitä [dataa] on koko ajan aktiivisesti puhdistettava ja ylläpidettävä. Se työ ei lakkaa koskaan.” (Tietojärj.P.)*

*”No se [datan laatu] vaatii kyllä ihan jatkuvaa ylläpitoa, koska meidän toiminta muuttuu... niin datan pitää heijastella koko ajan sitä, mitä meillä reaalielämässä tapahtuu ja sen takia erilaisen master datan ylläpito on tosi tärkeää.” (TalousP.)*

Kuvassa 6.3 on esitetty tietojärjestelmäkokonaisuuden tavoitetilanne tuotetiedon suhteen. Tietojärjestelmäkokonaisuuteen on lisätty ERP-järjestelmä. Excelistä on päästy täysin eroon.



**Kuva 6.3.** Kohdeyrityksen tietojärjestelmäkokonaisuuden tavoitetila.

Kuvassa 6.3 PLM- ja ERP-järjestelmät on integroitu niin, että käyttäjän tarvitsee luoda nimike ainoastaan PLM-järjestelmään. PLM-järjestelmästä nimike siirtyy automaattisesti ERP-järjestelmään, jossa sitä voidaan käyttää ERP-järjestelmän tarpeisiin.

## 7. POHDINTA JA JOHTOPÄÄTÖKSET

Vastauksena työn ensimmäiseen apukysymykseen (mikä rooli datan laadulla on tietojärjestelmä uudistuksissa ja mitä on master data) datan todettiin olevan yksi kolmesta merkittävästä tietojärjestelmä uudistuksen onnistumistekijästä, prosessi- ja käyttäjä sopivuuden ohella. Dataan ja sen laatuun on näin ollen todella tärkeää kiinnittää huomiota erityisesti tietojärjestelmä uudistuksen implementointivaiheessa. Työn tarkastelun kohteeksi nostettiin master data, koska datamigraatiossa siirretään pääasiassa master dataa. Master datan katsottiin olevan yksi neljästä keskeisestä datakategorista referenssidatan, transaktiodatan ja metadatan ohella. Sillä katsottiin olevan useita tyyppejä, kuten asiakas-, toimittaja- ja valmistajatiedot, sekä tässä työssä pääasiassa tarkasteltu tuotetieto.

Työn toiseen apukysymykseen (mistä master datan laatu koostuu) vastattiin kolmannessa luvussa, jossa todettiin master datan laadun koostuvan viidestä osatekijästä, jotka ovat tarkkuus, saatavuus, jatkuvuus, luotettavuus ja turvallisuus, sekä näiden neljästä näkökulmasta, jotka ovat data, tietojärjestelmä, käyttäjä sekä prosessi. Neljännessä luvussa tarkasteltiin kunkin osatekijän mittaamista, vastaten näin työn kolmanteen apukysymykseen (miten master datan laatua mitataan). Joidenkin osatekijöiden tarkastelu pystytään toteuttamaan kvantitatiivisesti, mutta useiden kartoittamiseen paras työkalu on käyttäjien haastattelu, joka tämän työn yhteydessä toteutettiin.

Näiden tietojen sekä haastattelutulosten pohjalta pystyttiin vastaamaan työn tutkimuskysymykseen (miten kohdeyrityksen master datan laadullisia ongelmia voidaan kartoittaa tietojärjestelmä uudistuksen yhteydessä ja miten niihin on reagoitava). Tätä on käsitelty luvussa 7.1. Muut työn tutkimuksessa esiin nousseet havainnot ja pohdinnat on esitetty luvussa 7.2.

### 7.1 Havaittujen ongelmien luokittelu

Datan tunteminen ja käyttäjien tarpeiden tiedostaminen ovat keskeisessä asemassa, jotta tietojärjestelmä kehitty nimenomaan kohdeyrityksen datan laatua ja dataprosesseja tukevaksi kokonaisuudeksi. Järjestelmän tuottaman datan laatu riippuu järjestelmän suunnittelusta (Wand & Wang 1996).

Järjestelmän suunnittelulla ei kuitenkaan pystytä suoraan vaikuttamaan siihen, miten dataa käytetään (Wand & Wang 1996). Datan ja tietojärjestelmien käyttäjien ottaminen mukaan datan laadun kehittämisprosessiin on tärkeää prosessin onnistumisen kannalta



(Foley & Helfert 2010). Tutkimuksen haastattelut toimivat osaltaan käyttäjiä osallistavina toimenpiteinä. Taulukossa 7.1 on esitetty tiivistettynä haastatteluissa esiin nousseet ongelmat, näiden ratkaisut sekä mahdolliset tulevaisuuden haasteet. Kustakin ongelma on tunnistettu myös ongelman näkökulma, joka ohjaa ongelmaan reagoimista. Taulukon sisältö on jaoteltu datan laadun osatekijöiden mukaan.

**Taulukko 7.1. Haastatteluissa esiin nousseet ongelmat, näiden ratkaisut ja lisähuomiot.**

|                     | Ongelmat  | Ratkaisut   | Lisähaasteet  | Näkökulma           |
|---------------------|---|---|---|---------------------|
| Nro                 | Tarkkuus  |   |   |                     |
| 1                   | Tyypiltään erilaiset nimikkeet samassa listassa sekoittuvat | Selkeä tietomalli, joka ohjaa/pakottaa tiedon täyttämiseen                        | Tietomalli, joka vastaa kaikkia tarpeita  | Data                |
| 2                   | Tuoterakenteen siirtäminen uuteen formaattiin               | Tuoterakenteeseen panostaminen heti alkuvaiheessa                                 | Uudessa järjestelmässä toimiva ja tavoitteita vastaava tuoterakenne             | Data                |
| 3                   | Data vastaa käyttäjien ja yrityksen todellisia tarpeita     | Jatkuva ylläpito ja huolehtiminen master datasta                                  | Joustavuus - järjestelmä mukautuu muuttuviin tarpeisiin                         | Prosessi            |
| 4                   | Datan tulkinnanvaraisuus                                    | Datan laatuun panostaminen  | Sama data sisäisissä toiminnoissa ja esim. tulevaisuuden verkkokaupassa         | Käyttäjä            |
| <b>Saatavuus</b>    |   |   |   |                     |
| 5                   | Yhteiskäyttö  | Järjestelmä estää samanaikaisen käytön  | Lukkoon jäävät nimikkeet estävät käytön   | Tietojärj.          |
| 6                   | CAD-tiedostot ja dokumentit kansiorakenteessa               | Data löytyy yhdestä paikasta  | Tietojärjestelmäriski   | Tietojärj.          |
| 7                   | Ei varsinaisia hakutoimintoja/-kenttiä                      | Selkeä tietomalli, joka mahdollistaa monipuolisen haun                            | Tietomalli, joka vastaa kaikkia tarpeita, myös hakua                            | Data/<br>Tietojärj. |
| 8                   | Datan liian suuri tai liian pieni määrä                     | Hakutulosten rajaaminen järjestelmässä. Kaikki tarvittava data järjestelmään      | Integraation rajoittaminen vain oleelliseen dataan                              | Tietojärj.          |
| <b>Jatkuvuus</b>    |   |   |   |                     |
| 9                   | Tietojen manuaalinen copy-paste järjestelmien välillä       | Järjestelmien välinen integraatio   | Integraation pysyminen ajantasalla  | Tietojärj.          |
| 10                  | Uuden tiedon päivittäminen moneen paikkaan                  | Järjestelmien välinen integraatio   |   | Tietojärj.          |
| 11                  | Datan tuottaminen järjestelmäkokonaisuuteen                 | Datalla on yksi master järjestelmä  | Järjestelmähierarkian säilyttäminen (ei ajauduta "kahden järjestelmän malliin") | Tietojärj.          |
| 12                  | Käyttäjät eivät ymmärrä tekojensa vaikutuksia               | Käyttäjien koulutus   |   | Käyttäjä            |
| 13                  | Datan sisällön yhtenäisen formaatin ja muodon puute         | Määritelty datan tuottamisen prosessi (tuottaa laadukasta dataa)                  | Kaikki data samassa formaatissa   | Data                |
| <b>Luotettavuus</b> |   |   |   |                     |
| 14                  | Inhimilliset virheet  | Määritelty datan validointiprosessi on joustava ja selkeä mutta myös standardoitu | Virheiden kulkeutuminen integraation kautta                                     | Data                |
| 15                  | Ei historiatietoja  | Järjestelmä hoitaa  |   | Tietojärj.          |
| 16                  | Muutosvastarinta  | Käyttäjien koulutus   | Motivoituneet käyttäjät   | Käyttäjä            |
| <b>Turvallisuus</b> |   |   |   |                     |
| 17                  | Hyvin rajalliset mahdollisuudet käytön rajoittamiseen       | Käyttöoikeuksien rajoittaminen tietojärjestelmässä                                | Tietojärjestelmäriski   | Tietojärj.          |
| 18                  | Varmuuskopioiminen hankalaa                                 | Data on helppo varmuuskopioida yhdestä sijainnista                                |   | Prosessi            |
| 19                  | Riippuvuus järjestelmän (ulkopuolisista) osaajista          | Tietojärjestelmävastaavien koulutus   |   | Käyttäjä            |

## Tarkkuus

Tarkkuuteen liittyviä ongelmia tunnistettiin neljä. Ensimmäinen ongelma oli erilaisten nimikkeiden käsitteleminen samassa listassa (ongelma 1), mikä oli yksi keskeinen tulkinanvaraisuuden aiheuttaja (ongelma 4). Ratkaisuna ongelmaan 1 on selkeän ja kattavan tietomallin luominen, joka vastaa laajalti tarpeisiin ja mahdollistaa erilaisten nimikkeiden luokittelun. Tämä on myös osaltaan ratkaisu ongelmaan 4.

Toinen merkittävä ongelma liittyi tuoterakenteeseen (ongelma 2). Aiemmin tuoterakennetta hallittiin osittain Excelissä, osittain CAD-mallissa. Haasteen muodostaa näiden formaattien yhdistäminen yhdeksi yhtenäiseksi rakenteeksi tietojärjestelmään.

Kolmas havaittu seikka ei ollut niinkään vielä ongelma, vaan ennemmin huoli. Datan on vastattava käyttäjien ja yrityksen todellisia tarpeita (ongelma 3). Tämä edellyttää jatkuvaa datan ylläpitoa ja prosessin muutosten seuraamista.

Tarkkuuteen liittyvistä ongelmista ongelmat 1 ja 2 kuuluvat datanäkökulmaan, joten nämä on korjattava ennen migraatiota tai sen yhteydessä. Muussa tapauksessa ongelmat siirtyvät uuteen tietojärjestelmään. Ongelmia 3 ja 4, jotka kuuluvat prosessi- ja käyttäjänäkökulmiin, ei voida suoraan korjata edeltä käsin, vaan niitä on pidettävä silmällä ja niihin on panostettava pitkällä aikavälillä.

## Saatavuus

Haastatteluissa nousi monesti esiin kohdeyrityksen dokumenttidatan yhteiskäytön (ongelma 5) ja saatavuuden (ongelmat 6 ja 7). Kohdeyrityksen uusi PLM-järjestelmä valvoo, ettei useampi käyttäjä pääse muokkaamaan samaa dataa samanaikaisesti. Tällöin vältytään aiemmin esiintyneiltä samanaikaisten muokkausten ongelmilta. Toisaalta järjestelmän mahdollistama nimikkeiden lukitseminen aiheuttaa myös jokapäiväisiä ongelmia lukkoon unohtuneiden nimikkeiden tapauksissa.

Dokumenttitiedostojen tallentaminen valtavaan kansiorakenteeseen aiheutti ennen merkittäviä haasteita (ongelma 6). Luokittelemalla dokumentit ja tallentamalla dokumentit PLM-järjestelmään niiden saatavuus parani valtavasti, kuten myös kaiken muun datan. Datan löytyminen yhdestä paikasta helpottaa merkittävästi datan hakemista, mutta toisaalta altistaa datan myös tietojärjestelmäriskille. Järjestelmän hajoaminen tarkoittaisi, ettei dataan ole enää pääsyä ja pahimmassa tapauksessa dataa menetettäisiin pysyvästi. Tietojärjestelmäriski on huomioitava ottamalla käyttöön järjestelmällinen varmuuskopiointikäytäntö.

Kun data sijaitsee yhdessä paikassa, dataan on helpompi kohdistaa hakutoimintoja. Vanhassa kansiorakenteessa vastaavia hakutoimintoja ei ollut (ongelma 7). Toimiva haku edellyttää selkeää tietomallia. Haasteen tämä aiheuttaa siinä, että saman tietomallin on toimittava monessa tilanteessa ja vastattava eri käyttäjien tarpeisiin, sillä kaikille erillisille tapauksille ei ole mielekästä rakentaa omaa tietomallia. Esimerkiksi kohdeyrityksen tuotekehityksen, valmistuksen ja varaston dokumentit olivat ennen osittain omissa tiedostosijainneissaan ja ne sisälsivät toisistaan eroavia tietoja. PLM-järjestelmässä yhden dokumenttitietomallin on pystyttävä vastaamaan näiden käyttäjien tarpeisiin.

Datan saatavuuteen yleisesti liittyy vahva käyttäjäsidonnaisuus, joka näkyi osittain myös tutkimuksessa. Käyttäjillä on datan suhteen omat tarpeensa, joiden täyttymisen perusteella datan saatavuutta arvioidaan. Tämän ilmiön ovat havainneet myös Strong et al. (1997). Vaikka datan hallinnoijien näkökulmasta kaikki data on helposti saatavilla tietojärjestelmästä, käyttäjät saattavat tarvita datan pohjalta muodostettua informaatiota. Datan hallinnoijien näkökulmasta datan saatavuus on siis hyvä, kun taas datan käyttäjien näkökulmasta saatavuus on huono. Tästä syystä datan hallinnoijien ja tietojärjestelmien kehittäjien on oleellista tietää, mihin ja miten dataa käytetään, jotta hakumahdollisuudet voidaan rakentaa mahdollisimman hyvin käyttäjien tarpeita palveleviksi.

Hyvät hakutoiminnot mahdollistavat hakutulosten tehokkaan rajaamisen, mikä vähentää suuren datamäärän aiheuttamia ongelmia (ongelma 8). Kohdeyrityksen datanimikkeiden kokoluokka on kirjoitushetkellä noin 6000 nimikettä, joista noin puolet ovat tuotenimikkeitä ja puolet dokumentteja. Kohdeyrityksen käyttämää tuotetiedon tietomallia kehitettiin tietojärjestelmä uudistuksen yhteydessä. Tietomallissa on kirjoitushetkellä 13 tietokenttää, joista 7 on määritelty pakollisiksi (tosin järjestelmä ei vielä kirjoitushetkellä pakota käyttäjää täyttämään näitä kenttiä). Kuten aiemmin on mainittu, tietomallin arvioiminen tietojärjestelmä uudistuksen yhteydessä voi hyvin olla harhaanjohtavaa, koska dataa joudutaan muun muassa muokkaamaan uuteen ympäristöön sopivaksi. Taulukossa 7.2 on esitetty suuntaa antavana tarkasteluna kohdeyrityksen tuotenimikkeiden määrä ja näistä ylimääräisten nimikkeiden osuudet sekä kohdeyrityksen tietomallin kenttien keskimääräiset täyttöasteet tammikuussa (noin kolme viikkoa PLM-järjestelmän käyttöönoton jälkeen) sekä elokuussa (noin seitsemän kuukautta PLM-järjestelmän käyttöönoton jälkeen). Muutokset on kuvattu kokonaismäärän osuuksina.

**Taulukko 7.2.** Kohdeyrityksen tietomallin tietokenttien keskimääräiset täyttöasteet.

|                                  | Tammikuu | Elokuu | Muutos |
|----------------------------------|----------|--------|--------|
| Tuotenimikkeiden määrä           | 2595     | 2924   | 12,7 % |
| Ylimääräisiä nimikkeitä          | 2,4 %    | 5,6 %  | 3,1 %  |
| Pakollisten kenttien täyttöaste  | 80,8 %   | 96,9 % | 16,1 % |
| Täydentävien kenttien täyttöaste | 26,1 %   | 38,5 % | 12,4 % |

Taulukosta 7.2 havaitaan, että tietokenttien täyttöasteet ovat parantuneet huomattavasti, mutta vielä on kehitettävää, erityisesti täydentävien tietokenttien (ei-pakolliset tietokentät) suhteen. Lisäksi ylimääräisten nimikkeiden osuuden kasvu viestii jostakin ongelmasta. Kasvun voidaan katsoa johtuneen kahdesta asiasta. Ensinnäkin uusi tietojärjestelmä teki ylimääräisten nimikkeiden tunnistamisesta helpompaa, jolloin niitä tunnistettiin vanhasta datasta enemmän. Toinen vaikuttava tekijä oli se, että käyttäjät eivät olleet vielä oppineet täysin tietojärjestelmän käyttöä, jolloin syntyy helpommin virhenimikkeitä. Esimerkiksi, jos käyttäjä ei löydä jotakin tiettyä nimikettä, hän saattaa tulkita, ettei sitä vielä ole järjestelmässä. Tällöin hän luo uuden nimikkeen, joka myöhemmin tunnistetaan toisen nimikkeen duplikaatiksi.

Saatavuuden ongelmien voidaan käytännössä kaikkien katsoa kuuluvan tietojärjestelmänäkökulmaan. Ainoa poikkeus on ongelma 7, johon liittyy myös datanäkökulma. Tämä ongelma on huomioitava tietomallia rakentaessa, joka esitettiin ratkaisuksi ongelmaan 1. Tietojärjestelmä uudistus tulee siis lähtökohtaisesti korjaamaan saatavuuden ongelmat, kunhan ne huomioidaan järjestelmää käyttöönottaessa ja myöhemmin sitä kehitettäessä.

### Jatkuvuus

Datan jatkuvuus oli kohdeyrityksessä merkittävä ongelma ennen tietojärjestelmä uudistusta. Datan hallitseminen useassa sijainnissa ja järjestelmässä ilman järjestelmien välistä integraatiota vaati manuaalista tiedon siirtämistä paikasta toiseen (ongelma 9), mikä puolestaan sai aikaan virheitä. Vastaavasti uuden tiedon tai muutoksen saattoi joutua syöttämään useampaan paikkaan (ongelma 10). Tietojärjestelmä uudistus poistaa lähtökohtaisesti nämä ongelmat, sillä uudessa tietojärjestelmäkokonaisuudessa järjestelmät on integroitu keskenään, eikä manuaalista tiedon siirtoa enää tarvita.

Toisaalta integraatio asettaa tarpeen määritellä kullekin datalle pääjärjestelmä, jossa sitä hallinnoidaan (ongelma 11). Pääjärjestelmästä data siirretään muihin järjestelmiin. Tällöin on tärkeää, että datan tuottamisen prosessi on hyvin käyttäjillä tiedossa (ongelma 12). Jos käyttäjä onnistuisi luomaan uuden nimikkeen järjestelmään, joka ei ole kyseisen nimikkeen pääjärjestelmä, nimike ei siirrykään muihin järjestelmiin. Tämä mahdollisuus rajataan pois käyttöoikeuksia rajaamalla, mutta ongelma säilyy silti käyttäjien hämmennyksenä. Ongelman ratkaisu on käyttäjien koulutus.

Jatkuvuuteen liittyen nousi esiin myös yksi datanäkökulman ongelma. Datan aikaisemman tietomalliin oli muodostettu hyvin vähän sääntöjä datan sisällön formaatista (ongelma 13). Esimerkkinä yksi käyttäjä saattoi luoda nimikkeen putkelle ja merkitsi sille halkaisijan millimetreissä. Toinen käyttäjä loi myös putken, mutta ilmoittikin halkaisijan tuumissa. Kolmas käyttäjä täytti samaan kenttään vain putken materiaalin, joka oli hänelle erityisen merkittävä tekijä. Datan tietomallin vapaasti täytettävillä kentillä on määriteltävä selkeät formaatit, joihin käyttäjät on opastettava. Näiden formaattien toteutumista on valvottava dataa validoidessa. Datanäkökulman ongelmana tämä on myös syytä korjata ennen datamigraatiota tai sen yhteydessä.

## **Luotettavuus**

Datan luotettavuuden keskeisin ongelma on inhimilliset virheet (ongelma 14). Virheiden syntymistä pystytään vähentämään rakentamalla datan validointiprosessi, jossa uusi data ja muokkaukset kulkevat aina validoinnin kautta. Yrityksen ja datamäärien ollessa vielä suhteellisen pieniä, datan validointi pystytään toteuttamaan melko suoraviivaisesti ja kattavasti. Maydanchik (2007, s26) arvioi, että kunkin yrityksen datasta 15-20 % muuttuu vuosittain. Näin ollen kohdeyrityksessä tulee vuosittain nimikemuutoksia noin 1200 (20 %), joka tarkoittaa kuukautta kohden 100 muutosta. Tämä määrä on vielä käytävissä läpi, mutta yrityksen kasvaessa ja datan lisääntyessä datan validointi tulee olemaan todella työlästä ja aikaa vievää, mikä saattaa muodostaa siitä pullonkaulan datan luomiseen ja muokkaamiseen. Nopea ratkaisu on kohdentaa enemmän henkilöstöä ja resursseja datan validointiin. Datamäärän kasvaessa edelleen saattaa tulla vastaan piste, jossa datan validointi on suoritettava pääasiassa datan laatusääntöjen avulla.

Ennen tietojärjestelmä uudistusta datasta jäi hyvin rajallinen määrä historiatietoa. Tämä muodosti ongelman, jos tuli tarve selvittää esimerkiksi mitä datassa on muutettu sekä kuka muutoksen on tehnyt ja miksi (ongelma 15). Tämä on ongelma, joka korjaantui erinomaisesti uuden tietojärjestelmän käyttöönotolla.

Uusiin tietojärjestelmiin liittyy aina enemmän tai vähemmän käyttäjien muutosvastarintaa (ongelma 16), joka johtuu usein järjestelmän käyttäjien ja ylemmän johdon tietojärjestelmän käyttöön liittyvien näkemysten välisestä ristiriidasta (Legare 2002) sekä epävarmuudesta uutta järjestelmää kohtaan (Hong & Kim 2002). Tähän ratkaisuna on käyttäjien osallistaminen, koulutus ja motivointi uusiin järjestelmiin osoittamalla, miten uudet järjestelmät helpottavat pitkällä aikavälillä heidän työtään. Myös jatkuvan palautteen antamisen mahdollistaminen alentaa muutosvastarintaa (Shaul & Tauber 2013).

## **Turvallisuus**

Ennen tietojärjestelmä uudistusta käyttäjien oikeuksien rajoittamisen mahdollisuudet olivat hyvin rajalliset (ongelma 17). Tämä johti ajoittain osaltaan virheisiin, joissa käyttäjä muokkasi epähuomiossa väärää nimikettä tai jopa vahingossa poisti dataa. Uusi tietojärjestelmä korjaa tämän ongelman, sillä tietojärjestelmässä käyttöoikeudet pystyy määrittelemään hyvin tarkasti.

Koska data oli ennen uudistusta säilötty hyvin laajaan kansiorakenteeseen, sen varmuuskopioiminen oli hankalaa (ongelma 18). Tämä oli merkittävä potentiaalinen riski. Uudistuksen jälkeen varmuuskopiointi on huomattavasti helpompaa, koska kaikki data sijaitsee yhdessä paikassa. Tämä helpottaa myös automaattisten, säännöllisten varmuuskopiointisääntöjen luomista.

Viimeinen ongelma (19) liittyy uusien tietojärjestelmien mahdollisesti muodostamiin riippuvuuksiin. Entiset datan käytön ja hallinnoinnin työkalut, kuten Excel, ei vaatinut ulkopuolista osaamista. Uudet tietojärjestelmät sen sijaan ovat huomattavasti monimutkaisempia hallinnoida, jolloin osa osaamisesta jää helposti organisaation ulkopuolelle yhteistyökumppaneille. Käyttäjien huoli tällaisista riippuvuuksista johtuu siitä, että näillä yhteistyökumppaneilla on paljon asiakkaita, ja pientäkin muutosta saatetaan joutua odottamaan kauan, jos sitä ei osata tehdä itse. Ratkaisuna tähän ongelmaan on kustannustehokkuuden puitteissa hankkia sopiva määrä järjestelmäosaamista organisaation sisälle.

Mikään ei sulje pois sitä mahdollisuutta, etteikö tulevaisuudessa tulisi uusia datan laatuun liittyviä ongelmia tai etteivätkö uudet tietojärjestelmät toisi tullessaan uusia haasteita. Keskittymällä kuitenkin reagoimaan ja korjaamaan nämä havaitut ongelmat, tietojärjestelmä uudistuksen voidaan uskoa sujuvan sulavasti ainakin datalähtöisen sopivuuden osalta.

## 7.2 Yleiset havainnot tutkimuksesta

Datan laadun kokonaisuuden tunteminen on todella tärkeää dataa ja tietojärjestelmiä kehitettäessä. Yksi työn keskeisimmistä havainnoista oli datan laadun kontekstisidonnaisuus. Data sijaitsee aina jossakin ja siihen päästään käsiksi jollakin järjestelmällä. Koska datan ja käyttäjän välissä on aina jokin järjestelmä, datan laatu saattaa näyttäytyä hyvin eri tavoin käytettäessä eri järjestelmiä. Tämä korostui erityisesti kohdeyrityksen tapauksessa, jossa vanhaa dataa alettiin käyttämään uudella järjestelmällä.

Esimerkiksi käyttäjällä voi olla tallennettuna tietokantaan lista asiakkaista, joka sisältää täydellisesti kaiken tarvittavan tiedon. Jos käytettävissä on kuitenkin järjestelmä, joka mahdollistaa haun pelkän asiakasnumeron perusteella, datan saatavuus on todella huono. Tällöin olisi osittain harhaanjohtavaa puhua varsinaisen datan laadun ongelmasta, koska itse datassa ei ole mitään vikaa, vaan kyseessä on tietojärjestelmän puute. Toki useissa tutkimuksissa, kuten Strong et al. (1997) ja Wang (1998), esimerkiksi saatavuus oli luokiteltu omaksi kategoriakseen, mutta rajanveto eri näkökulmien välillä pitäisi olla vielä selkeämpi.

Datan kontekstisidonnaisuuden ja riippuvuuden käytetyistä tietojärjestelmistä, prosesseista ja ihmisistä on havainnut myös esimerkiksi Vilminko-Heikkinen (2017) väitöskirjassaan. Datan laadusta ei voi puhua puhumatta samalla tietojärjestelmistä, prosesseista ja dataa käsittelevistä ihmisistä, kuten on jo monesti käynyt ilmi. Kaikki nämä on otettava huomioon datan laatua kehitettäessä.

Haug & Stentoft (2011) tuovat esiin, että datan puutteellinen hallinta johtaa organisaation eri toimintojen ”siiloutumiseen”. Datan suhteen tämä tarkoittaa sitä, että kullakin organisaation toiminnolla on oma datansa ja omat käytännöt sen hallinnoimiseen (Fatehali 2011). Työn tutkimuksessa tämä havaittiin selkeästi. Kohdeyrityksen eri toiminnoilla oli paikoin erilaiset käsitykset siitä, mitä hyvälaatuiselta datalta vaaditaan. Nämä ristiriitaiset käsitykset nousivat selvästi esille, kun kohdeyritys siirtyi kunkin toiminnon omasta Excel-tilukosta hallinnoimaan dataa yhdessä tietojärjestelmässä. Tällöin yhden datan odotettiin vastaavan usean toiminnon tarpeisiin, mikä aiheutti haasteita datan hallinnassa. Silvola (2018) puhuu tällaisen yhden yhteisen master datan puolesta. On selvää, että tämän toteuttamisessa on valtavasti haasteita, mutta tämän työn tutkimuksen perusteella se vaikuttaa mahdolliselta.

Haastatteluista kävi ilmi, että yksilöiden näkemyksissä datan laadusta korostuvat usein vain tietyt näkökulmat. Näin ollen datan laatua tarkastellessa ja kartoittaessa on syytä hyödyntää useiden eri toimintojen henkilöstöä. Organisaatioissa, joissa datan käyttö ja

tietojärjestelmät ovat edenneet stabilisointivaiheesta pitkälle käyttövaiheeseen, on myös syytä hyödyntää datan laadun kvantitatiivista analysointia, joka esitettiin luvussa 4.

Mielenkiintoisena huomiona datan luotettavuuden käyttäjänäkökulman eli datan maineen ja uskottavuuden kartoitukseen oli, että optimi ei välttämättä ole täysi luotto dataan. Maineen ja uskottavuuden kartoituskeinoksi esitettiin kyselyä, jossa voidaan kysyä esimerkiksi luottoa dataa kohtaan asteikolla 1-10. Kuten muun muassa Redman (2008) mainitsee, tietojärjestelmien uskotaan monesti automaattisesti korjaavan dataan liittyviä ongelmia. Täysi 10/10 luottamus dataa kohtaan saattaa viestiä sopivan kriittisen lähestymistavan puutteesta. Tällöin on riskinä, että datassa olevat virheet pääsevät leviämään nopeammin ja esimerkiksi kulkeutumaan herkemmin yhteistyökumppaneille ja asiakkaille. Näin ollen optimi datan maineelle ja uskottavuudelle saattaakin olla ennemmin 8 tai 9 asteikolla 1-10.

Erityisesti joidenkin datan käyttäjien, kuten talouspäällikön, haastatteluissa nousi esiin tunnettu ilmiö, jossa dataan ja datan laatuun kiinnitetään huomiota vasta kun ongelmia alkaa esiintyä. Muun muassa Russom (2006) on havainnut vastaavaa. Datan laadun suhteen on syytä pyrkiä reaktiivisesta lähestymistavasta proaktiiviseen, ennaltaehkäisevään lähestymistapaan (Russom 2006). Tutkimus on myös osoittanut, että yleisin ongelma datan heikosta laadusta on ylimääräinen työhön kuluva aika (Russom 2006), mikä nousi useasti esiin myös tämän työn haastatteluissa.

### 7.3 Teoreettiset kontribuutiot

Työssä muodostettu datan laadun viitekehys helpotti huomattavasti datan laadun kartoittamista hyödyntämällä neljää näkökulmaa. Verrattuna aiempiin tutkimuksiin, kuten Levitin & Redman (1995), Wand & Wang (1996), Strong et al. (1997), Tayi & Ballou (1998), Wang (1998), Lee et al. (2002), Pipino et al. (2002) sekä Zhang et al. (2019), viitekehys muodosti merkittävästi helpommin hallittavan ja hahmotettavan kokonaisuuden datan laadusta. On myös huomattava, että työssä muodostettu viitekehys on kokonaisuutena laajempi näkemys datan laadusta kuin useimmissa aiemmissä tutkimuksissa käytetty tarkastelu. Esimerkiksi Wang (1998) jakaa datan laadun tekijät neljään kategoriaan, jotka ovat sisäiset, saatavuudelliset, kontekstuaaliset sekä esitysmuotoon liittyvät tekijät. Tämän työn viitekehyyksen neljä näkökulmaa (data, tietojärjestelmä, käyttäjä ja prosessi) olivat datan laadun tekijöiden jaotteluperusteina paremmin tasapainossa kuin Wangin (1998) käyttämät neljä kategoriaa, joiden sisältämien tekijöiden määrä vaihteli kahdesta viiteen. Koska tämän työn haastatteluissa kunkin osatekijän jokainen näkö-



kulma nousi haastatteluissa esiin useaan kertaan, on mahdollista, että aiemmissa tutkimuksissa, joissa ei ole käytetty yhtä montaa osatekijää, on jäänyt jokin datan laadun aspekti vähäiselle huomiolle.

Datan laadun ongelmien luokitteluun näkökulmien mukaan data-, tietojärjestelmä-, käyttäjä sekä prosessiongelmien osoittautui tärkeäksi ja on näin ollen työn toinen teoreettinen kontribuutio. Tärkeäksi luokittelun tekee se, että eri näkökulmien ongelmiin on reagoitava eri tavoin. Vastaavaa luokittelua ei vaikuta löytyvän alan kirjallisuudesta. Joissakin aiemmissa tutkimuksissa, kuten Pipino et al. (2002) ja Zhang et al. (2019) ei käytetä lainkaan tekijöiden luokittelua. Tämän luokittelun tärkeys korostuu erityisesti tietojärjestelmän implementointivaiheessa, jossa dataa valmistellaan migraatioon. Implementoinnin on lähtökohtaisesti esimerkiksi korjattava tietojärjestelmän näkökulman ongelmat. Jos näin ei tapahdu, uuden tietojärjestelmän merkitys vähenee huomattavasti.

Työn kolmantena teoreettisena kontribuutiona voidaan katsoa olevan datan laadun korostaminen tietojärjestelmä uudistuksen yhteydessä. Organisatorista sopivuutta käsittelevät tutkimukset keskittyvät usein tarkastelemaan organisaation prosessien mukautumista tai järjestelmän mukautumista kokonaisuutena (Hong & Kim 2002, Chen et al. 2009). Kenties data-lähtöinen sopivuus ei ole tärkein tietojärjestelmän implementoinnin onnistumisen tekijöistä, mutta se on ehdottomasti huomion arvoinen, mikä tuli ilmi myös tämän työn kohdeyrityksen tapauksessa.

## 7.4 Käytännön kontribuutiot

Työn sisältö helpottaa organisaatioita huomattavasti datan ja sen laadun huomioimisessa tietojärjestelmä uudistuksen yhteydessä ja myös muulloin. Viitekehys auttaa datan laadun kokonaisvaltaisessa hahmottamisessa ja mahdollistaa havaittujen ongelmien kartoittamisen taulukon 7.1 tapaan. Kartoitus selkeyttää ongelmiin reagoimista ja erityisesti organisaatioissa, joissa ei ole käytössä suuria resursseja, se auttaa priorisoimaan ongelmia.

Parhaassa tapauksessa työ myös osoittaa organisaatioille, että datan laatu ei kuvaa pelkästään tietojärjestelmässä olevaa tietomassaa. Siinä missä data ja tietojärjestelmä kattavat usein koko organisaation, myös datan laatu näyttäytyy kaikkialla ja vaikuttaa joko suoraan tai välillisesti kaikkiin organisaation toimintoihin. Datan laatu on paitsi tietomasojen objektiivista arvoa, myös ihmisten subjektiivista arvoa. Paraskaan data ei näyttyä laadukkaana, jos se ei palvele käyttötarkoitustaan, eikä paras tietojärjestelmä toimivana, jos käyttäjät eivät sitä hallitse. Näin ollen datan laadun hallinta on ajateltava ja toteutettava hyvin laajasti.

## 7.5 Tutkimuksen rajoitteet

Luvussa 5.1 esiteltiin kvalitatiivisen tutkimuksen viisi laatukriteeriä, jotka olivat uskottavuus, siirrettävyys, luotettavuus, varmennettavuus ja refleksiivisyys. Yksi keskeinen tutkimuksen uskottavuuden parantamisen työkalu on ”kolmiomittaus”, jossa pyritään hyödyntämään eri tietolähteitä saman tiedon saamiseksi (Korstjens & Moser 2018). Jos eri tyyppiset tietolähteet tuottavat saman tiedon, tiedon uskottavuus paranee. Työn tutkimuksessa haastateltiin kohdeyrityksen henkilöstöä eri toiminnoista. Lisäksi haastateltavissa oli datan tuottajia, käyttäjiä sekä hallinnoijia. Tutkimuksen tuloksena saatiin lista kohdeyrityksen datan laadun ongelmista. Listalla olevat ongelmat toistuivat useasti haastatteluissa. Uskottavuuden voidaan näin ollen katsoa toteutuneen tutkimuksessa hyvin.

Toinen laatukriteeri oli tutkimuksen siirrettävyys. Koska tutkimus oli yhden tapauksen tapaustutkimus, saadut tulokset ovat hyvin kontekstisidonnaisia. Tutkimukseen muodostettu datan laadun viitekehys sekä lähestymistapa datan laatua kohtaan ovat yleispäteviä, joten niitä voidaan todennäköisesti hyödyntää ja tutkia muissakin konteksteissa. Sen sijaan tutkimuksen pohjalta muodostettu listaus datan laadun ongelmista on vahvasti kontekstisidonnainen. Toki tunnistettuja ongelmia voi pyrkiä tunnistamaan toisesta vastaavassa tilanteesta olevasta organisaatiosta, mutta tämä ei ollut listauksen ensisijainen tarkoitus. Siirrettävyys toteutui siis osittain.

Tulosten luotettavuus ja varmennettavuus olivat kolmas ja neljäs laatukriteeri. Työkalu näiden kriteerien parantamiseen on ”jäljitysketju”, jossa kuvataan mahdollisimman läpinäkyvästi tutkimuksen toteutus ja eteneminen (Korstjens & Moser 2018). Kuvaukset tutkimuksesta, menetelmästä sekä tulosten analysoinnista löytyvät työn luvusta 5. Haastatteluista muodostettu listaus tuloksista on nykyhetken ongelmien kannalta sidoksissa ajanhetkeen. On mahdollista, että haastatteluissa ei tullut ilmi kaikkia organisaatiossa esiintyviä datan laadun ongelmia, koska sekä haastateltavien määrä että tarkasteltavan datan tyyppi oli rajattu. Haastatteluissa oli selvästi havaittavissa paikoin myös sidoksisuutta ajankohtaan, sillä haastattelijat toivat eniten esille viime päivinä ja viikkoina esillä olleita asioita ja ongelmia. Lisäksi uusia datan laadun ongelmia tulee todennäköisesti ilmenemään uusien tietojärjestelmien käytön vakiintumisen myötä. Näin ollen voidaan olettaa, että tutkimuksen toistaminen muuttaisi tuloksia jossain määrin. Kuten aiemmin mainittu, muodostettu viitekehys ja ongelmien kartoitustapa, joiden toimivuutta tutkimuksella myös pyrittiin testaamaan, ovat yleispäteviä ja niitä koskevat tulokset säilyvät todennäköisesti ennallaan. Luotettavuuden ja varmennettavuuden voidaan siis todeta myös toteutuneen osittain.

Viimeinen laatukriteeri oli refleksiivisyys. Korstjens & Moser (2018) nostavat tämän laatukriteerin työkaluksi päiväkirjan, johon tutkija kirjaa tutkimuksen aikana esiin nousseita huomioita, tunnistettuja oletuksia, ennako-odotuksia ja omia arvojaan. Tutkimuksen aikana ei pidetty päiväkirjaa. Ongelmalistan suhteen ei kuitenkaan ole syytä olettaa, että kirjoittajan subjektiivisuus olisi vaikuttanut listaan merkittävästi, sillä useat haastateltavat nostivat esiin itsenäisesti samat ongelmat. Toisaalta tämä voisi olla merkki siitä, että haastattelija johdatteli haastateltavat samoihin ongelmiin. Ei ole syytä uskoa näin tapahtuneen, sillä haastattelukysymykset pyrittiin pitämään mahdollisimman yleisellä tasolla ja kukin haastateltava nosti esiin oman näkökulmansa seikkoja. Sen sijaan datan laadun viitekehyksen ja lähestymistavan suhteen on täysin mahdollista, että tuloksien tulkinassa esiintyy jossain määrin tutkimuksen tekijän subjektiivisuutta. Esimerkiksi kunkin kvalitatiivisen ongelman luokitteluun johonkin työssä määritellyistä luokista on osoittain aina luokittelijan näkemys. Vastaavaa on havaittavissa myös siinä, että aiemmissa tutkimuksissa datan laadun osatekijöiden luokitteluissa on tutkimuskohtaisia eroja, kuten taulukko 3.2 osoitti. Toinen tutkija olisi esimerkiksi saattanut luokitella ongelmat eritavoin ja todennut, että muodostettu viitekehys on puutteellinen.

## 7.6 Tulevaisuuden tutkimusaiheet

Eppler (2003, s349-352) on myös muodostanut datan laadun viitekehyksen. Siinä missä tämä työ luokittelee datan laadun osatekijät neljään näkökulmaan, Eppler hyödyntää informaation arvoketjua ja jaottelee osatekijät dataongelmiksi, informaatio-ongelmiksi, kommunikaation ongelmiksi ja tietämyksen ongelmiksi. Vaikka Epplerin lähestymistapa on erilainen, viitekehyksissä on huomattavia yhtäläisyyksiä. Näiden viitekehysten tarkempi vertailu saattaisi avata lisää näkemyksiä datan laadusta.

Tässä työssä muodostettu datan laadun viitekehys toimi haastattelujen perusteella hyvin. Taulukossa 6.1 esitettiin haastatteluissa esiin nousseet datan laadun osatekijät esiintymismäärinä viitekehyksessä. Eri tekijöiden ja eri näkökulmien yhteenlasketut esiintymismäärät (vaakarivien summat ja pystysarakkeiden summat) olivat melko lähellä toisiaan välillä 24-41 esiintymiskertaa. Tämä viittaa siihen, että osatekijät ovat hyvin tasapainossa keskenään eli osatekijät on valittu oikein. Sama pätee myös näkökulmiin. Toisaalta on huomattava, että työn konteksti on saattanut ohjata haastattelujen kulkua tiettyyn suuntaan. Haastatteluissa ei kuitenkaan noussut esiin yhtäkään datan laadun osatekijää, jota ei olisi voinut esittää työssä muodostetun viitekehyksen avulla. Kuten tutkimuksen rajoitteissa mainittiin, on täysin mahdollista, että viitekehyksen arviointiin liittyy subjektiivisuutta. Viitekehyksen toimivuutta onkin siis syytä tutkia toisessa kontekstissa ja toisen tutkijan toimesta.

Taulukosta 6.1 havaittiin myös, että eri osatekijät painottuvat tiettyihin näkökulmiin. Tarkkuus painottuu dataan ja tietojärjestelmiin, saatavuus käyttäjään, jatkuvuus ja luotettavuus tietojärjestelmiin ja turvallisuus dataan ja tietojärjestelmiin. On hyvin mahdollista, että havaittu painottuminen saattaa olla täysin organisaatiokohtaista, osoittaen tarkasteltavan organisaation kriittiset kehityskohteet. Toisaalta, jos painottuminen ei ole organisaatiokohtaista, tämä saattaa osoittaa nämä osatekijöiden näkökulmat erityisen kriittiksi. Tieto kriittisyydestä helpottaisi datan laadun ongelmien ja kartoittamisen priorisointia erityisesti organisaatioissa, joissa on hyvin rajalliset resurssit.

Työn tutkimuksessa tarkasteltava datan tyyppi oli rajattu master dataan ja vielä edelleen tuotedataan. Vaikka tutkimuksessa sivuutettiin myös muun muassa dokumenttidataa, on mahdollista, että erilaisilla datatyypeillä on hieman erilaiset laatuksiteerit. Toisaalta työssä muodostettu viitekehys rakennettiin kirjallisuuden pohjalta, jolloin lähtökohtaisesti datan osatekijöissä ei pitäisi esiintyä erityisen paljon vinoumia tiettyihin datatyyppeihin. Viitekehysten ja lähestymistavan toimivuutta on kuitenkin tutkittava myös muiden datatyyppien tapauksissa.

## 7.7 Loppusanat

Datan laadun kehittäminen ja ylläpito on jatkuva prosessi, johon koko yrityksen on panostettava. Data on yrityksen toiminnan kannalta avainasemassa, minkä vuoksi puutteet datan laadussa heijastuvat ennemmin tai myöhemmin yrityksen liiketoimintaan. Kuten luvun 3 kuvasta 3.1 kävi ilmi, datan laatua ei voida ylläpitää huomioimatta kaikkia neljää näkökulmaa. Yhdellä hetkellä hyvinkin arvokas data saattaa liiketoimintastrategian muutoksen seurauksena muuttua taakaksi, jonka ylläpito on ainoastaan lisärasite ja ylimääräinen kustannus organisaatiolle. Juuri datan laadun kontekstisidonnaisuus tekee sen tarkastelusta vaikeaa. Työn tutkimuksessakin havaittiin useaan otteeseen, että erilaiset käyttäjät kokevat datan laadukkuuden hyvin eri tavoin. Datan laadun onkin lähdettävä datan käyttötarkoituksesta.

Tietojärjestelmien käyttöönotossa on hyvä tiedostaa, että jopa joustavimmatkin järjestelmät, kuten Umble et al. (2003) huomauttavat, tuovat aina organisaatioon oman logiikkansa ja omat muutoksensa. Aiempien tutkimusten lisäksi tässä työssä kerätty kokemus tietojärjestelmä uudistuksista osoittaa, että yritykset pääsevät monesti paljon helpomalla omaksumalla tietojärjestelmän tavan tehdä asioita kuin pakottamalla tietojärjestelmän noudattamaan yrityksen vanhaa käytäntöä.

Kohdeyrityksen tietojärjestelmä uudistuksen voidaan katsoa onnistuneen PLM-järjestelmän osalta. Kirjoitushetkellä myös ERP-järjestelmän käyttöönotto näyttää hyvältä. Pannostaminen datan laatuun läpi koko prosessin helpotti ja nopeutti huomattavasti käyttöönottoa. Uuden tietojärjestelmän kautta laadukkaana näyttäytyvä data motivoi selvästi myös käyttäjiä uuden järjestelmän käyttöön, vähentäen muutosvastarintaa. Kokonaisuutena voidaan katsoa, että työssä onnistuttiin vastaamaan tutkimuskysymyksiin sekä teoreettisesta että käytännön näkökulmasta.

## LÄHTEET

- Ballou, D. & Pazer, H. (1985). Modeling data and process quality in multi-input, multioutput information systems. *Management Science*, Vol. 31(2): 150–162.
- Berson, A. & Dubov, L. (2007). *Master Data Management and Customer Data Integration for a Global Enterprise*. McGraw Hill Professional.
- Bhatti, T. (2014). Critical Success Factors for the Acquisition of Enterprise Systems: Empirical Validation. *International Conference on Information Resources Management*.
- Breuer, T. (2009). Data quality is everyone's business—designing quality into your data warehouse—Part 1. *Journal of Direct, Data and Digital Marketing Practice*, Vol. 11(1): 20–29.
- Buffington, J. (2011). Comparison of mass customization and generative customization in mass markets. *Industrial Management & Data Systems*, Vol. 111(1): 41–62.
- Chen, H., Chen, S. & Tsai, L. (2009). A Study of Successful ERP – From the Organization Fit Perspective. *Systemics, Cybernetics and Informatics*, Vol. 7(4): 8–16.
- Chrysosolouris, G., Mavrikios, D., Papakostas, N., Mourtzis, D., Michalos, G. & Georgoulas, K. (2009). Digital manufacturing: history, perspectives, and outlook. *Proceedings of the Institution of Mechanical Engineers, Part B, Journal of Engineering Manufacture*, Vol. 223.
- Eppler, M. (2003). *Managing Information Quality*, 2nd edition. Springer-Verlag Berlin Heidelberg.
- Esteves, J. & Pastor, J. (2005). A Critical Success Factor's Relevance Model for SAP Implementation Projects. *Lähteestä Lau, L. (2005). Managing Business with SAP: Planning, Implementation, and Evaluation*. Idea Group Publishing, 240–261.
- Fatehali, M. (2011). Building the Business Case for Master Data Management in the Public Sector. An Oracle White Paper.
- Fichman, R., Kohli, R. & Krishnan, R. (2011). Editorial Overview—The Role of Information Systems in Healthcare: Current Research and Future Trends. *Information Systems Research*, Vol. 22(3).
- Foley, O. & Helfert, M. (2010). Information Quality and Accessibility. *Lähteestä Sobh, T. (2010). Innovations and Advances in Computer Sciences and Engineering*. Springer, 477–481.

Gerow, J., Thatcher, J. & Grover, V. (2015). Six Types of IT-Business Alignment: An investigation of constructs and their measurement. *European Journal of Information Systems*, Vol. 24(5): 465–491.

Hannus, J. (2004). Strategisen menestyksen avaimet: Tehokkaat strategiat, kyvykkyydet ja toimintamallit. ProTalent, Gummerus Jyväskylä.

Haug, A. & Stentoft, J. (2011). Barriers to master data quality. *Journal of Enterprise Information Management*, Vol. 24(3): 288–303.

Holland, C. & Light, B. (1999). A critical success factors model for ERP implementation. *IEEE Software*, Vol. 16(3): 30–36.

Hong, K. & Kim, Y. (2002). The critical success factors for ERP implementation: an organizational fit perspective. *Information & Management*, Vol. 40: 25–40.

Knolmayer, G. & Röthlin, M. (2006). Quality of material master data and its effect on the usefulness of distributed ERP systems. *Lecture Notes in Computer Science*, Vol. 4231: 362–371.

Korstjens, I. & Moser, A. (2018). Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing. *European Journal of General Practice*, Vol. 24(1): 120–124.

Lee, H., Ahn, H., Kim, J. & Park, S. (2006). Capturing and reusing knowledge in engineering change management: A case of automobile development. *Information Systems Frontiers*, Vol. 8(5): 375–394.

Lee, Y., Madnick, S., Wang, R., Wang, F. & Zhang, H. (2014). A Cubic Framework for the Chief Data Officer: Succeeding in a World of Big Data. *University of Minnesota, MIS Quarterly Executive*, Vol 13(1): 1–13.

Lee, Y., Strong, D., Kahn, B. & Wang, R. (2002). AIMQ: a methodology for information quality assessment. *Information & Management*, Vol. 40(2): 133–146.

Legare, T. (2002). The role of organizational factors in realizing ERP benefits. *Information Systems Management*, Vol. 19(4): 21–42.

Levitin, A. & Redman, T. (1995). Quality Dimensions of a Conceptual View. *Information Processing & Management*, Vol. 31(1): 81–88.

Lincoln, Y. & Cuba, E. (1985). *Naturalistic Inquiry*. Sage Publications.

Loshin, D. (2009). *Master Data Management*. Morgan Kaufmann.

Loshin, D. (2011). *The Practitioner's Guide to Data Quality Improvement*. Morgan Kaufmann.

- Martio, A. (2015). Tuotekonfigurointi ja tuotetiedon hallinta. Amartekno.
- Maydanchik, A. (2007). Data Quality Assessment. Technics Publications, LLC.
- McGilvray, D. (2008). Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information. Morgan Kaufmann Publishers.
- Meredith, J. (1998). Building operations management theory through case and field research. *Journal of Operations Management*, Vol. 16: 441–454.
- Olson, J. (2003). Data Quality – The Accuracy Dimension. Morgan Kaufmann.
- Ouertani, M., Baïna, S., Gzara, L. & Moreld, G. (2011). Traceability and management of dispersed product knowledge during design and manufacturing. *Computer-Aided Design*, Vol. 43(5): 546–562.
- Panian, Z. (2010). Some practical experiences in data governance. *World Academy of Science, Engineering and Technology*, Vol. 38: 150–157.
- Parr, A., Shanks, G. & Darke, P. (1999). Identification of necessary factors for successful implementation of ERP systems. Lähdeestä Ngwenyama, O., Introna, L., Myers, M. & DeGross, J. (1999). New information technologies in organizational processes: Field studies and theoretical reflections on the future of work, 99–119.
- Pipino, L., Lee, Y. & Wang, R. (2002). Data Quality Assessment. *Communications of the ACM*, Vol. 45(4): 211–218.
- Preston, D. & Karahanna, E. (2009). Antecedents of IS Strategic Alignment: A nomological network. *Information Systems Research*, Vol. 20(2): 159–179.
- Redman, T. (1998). The Impact of Poor Data Quality on the Typical Enterprise. *Communications of the ACM*, Vol. 41(2): 79–82.
- Redman, T. (2008). Data driven: profiting from your most important business asset.
- Reynolds, P. & Yeatton, P. (2015). Aligning Business and IT-strategies in multi-business organizations. *Journal of Information Technology*, Vol. 30: 101–118.
- Russom, P. (2006). Taking data quality to the enterprise through data governance. TDWI Report Series, The Data Warehousing Institute, Seattle.
- Sebastian-Coleman, L. (2013). Measuring Data Quality for Ongoing Improvement, 1<sup>st</sup> Edition. Morgan Kaufmann.
- Setia, P., Venkatesh, V. & Joglekar, S. (2013). Leveraging digital technologies: how information quality leads to localized capabilities and customer service performance. *MIS Quarterly*, Vol. 37(2): 565–A4.



- Shaul, L. & Tauber, D. (2013). Critical success factors in enterprise resource planning systems: Review of the last decade. *ACM Computing Surveys (CSUR)*, Vol. 45(4): 1–39.
- Silver, N. (2012). *The Signal and the Noise: Why So Many Predictions Fail - But Some Don't*. Penguin Press.
- Silvola, R. (2018). *One Product Data for Integrated Business Processes*. Väitöskirja, Oulun yliopisto.
- Silvola, R., Tolonen, A., Harkonen, J., Haapasalo, H. & Männistö, T. (2018). Defining one product data for a product. *International Journal of Business Information Systems*.
- Somers, T. & Nelson, K. (2001). The Impact of Critical Success Factors across the Stages of Enterprise Resource Planning Implementations. *Proceedings of the 34th Hawaii International Conference on System Sciences*.
- Strong, D., Lee, Y. & Wang, R. (1997). Data Quality in Context. *Communications of the ACM*, Vol. 40(5): 103–110.
- Tallon, P. (2008). A Process-Oriented Perspective on the Alignment of Information Technology and Business Strategy. *Journal of Management Information Systems*, Vol. 24(3): 227–268.
- Tayi, G. & Ballou, D. (1998). Examining Data Quality. *Communication of the ACM*, Vol. 41(2): 54–57.
- Umble, E., Haft, R. & Umble, M. (2003). Enterprise resource planning: Implementation procedures and critical success factors. *European Journal of Operational Research*, Vol. 146: 241–257.
- Vilminko-Heikkinen, R. & Pekkola, S. (2012). Organizational Issues in Establishing Master Data Management Function. *Lähteestä Proceedings of the 17th International Conference on Information Quality, ICIQ 2012, 16-17 November 2012, Paris, France*.
- Vilminko-Heikkinen, R. (2017). *Data, Technology, and People: Demystifying Master Data Management*. Väitöskirja, Tampereen teknillinen yliopisto, julkaisu 1457.
- Voss, C., Tsikriktsis, N. & Frohlich, M. (2002). Case research in operations management. *International Journal of Operations & Production Management*, Vol. 22(2): 195–219.
- Wand, Y. & Wang, R. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, Vol. 39(11): 86–95.
- Wang, R. (1998). A Product Perspective on Total Data Quality Management. *Communications of the ACM*, Vol. 41(2): 58–65.
- Watts, S., Shankaranarayanan, G. & Even, A. (2009). Data quality assessment in context: A cognitive perspective. *Decision Support Systems*, Vol. 48(1): 202–211.

Yayla, A. & Hu, Q. (2012). The Impact of IT-Business Strategic Alignment on Firm Performance in a Developing Country Setting: Exploring moderating roles of environmental uncertainty and strategic orientation. *European Journal of Information Systems*, Vol. 21(4): 373–387.

Yin, R. (2008). *Case Study Research: Design and Methods*, 4th Edition. Applied Social Research Methods. Sage Publications.

Zhang, R., Indulska, M. & Sadiq, S. (2019). Discovering Data Quality Problems. *Business & Information Systems Engineering*, Vol. 61(5): 575–593.

Zwicker, R. & de Souza, C. (2005). *SAP R/3 Implementation Approaches: A Study in Brazilian Companies*. Lähteestä Lau, L. (2005). *Managing Business with SAP: Planning, Implementation, and Evaluation*. Idea Group Publishing, 198–221.

## LIITE A: HAASTATTELUTUTKIMUKSEN KYSYMYSRUNKO

- Ketä haastatellaan
- Päivämäärä ja kellonaika

\*Lyhyt kuvaus diplomityöstä ja haastattelun tavoitteista

- Mitä datan laatu merkitsee sinulle työssäsi?
- Mikä on sinun mielestäsi hyvälaatuista dataa? Mitä
- Millaisia datan laadun keskeisiä tekijöitä sinulle tulee mieleen?

Ennen tietojärjestelmä uudistusta

- Miten kuvailisit yrityksen tuotehallinnan ”tietojärjestelmää” ennen uudistusta?
- Miten kuvailisit datan laatua ennen uudistusta?
  - o Mitä hyvää datan laadussa oli?
  - o Mitä haasteita datan laadussa oli?

Tietojärjestelmä uudistuksen käyttöönottoaikana

- Miten koet tietojärjestelmän (PLM) ja sen roolin yrityksessä tässä vaiheessa?
- Miten kuvailisit datan laatua uudistuksen aikana (PLM otettu käyttöön, ERP tulossa)
  - o Mitä hyvää datan laadussa on?
  - o Mitä haasteita datan laadussa on?

Tulevaisuudessa

- Mitä tietojärjestelmän (ERP + PLM) pitää mahdollistaa tulevaisuudessa?
- Miten näet datan laadun kehittyvän tulevaisuudessa?
  - o Mitä mahdollisuuksia?
  - o Mitä riskejä?
- Millaisilla toimenpiteillä datan laatu saadaan pysymään hyvänä?
- Millaisia uusia vaatimuksia näet kohdistuvan datan laatuun seuraavan 5 vuoden aikana?
- Onko vielä jotain ajatuksia mitä tietojärjestelmä uudistus herättää?
- Onko vielä jotain ajatuksia mitä datan laatu herättää?